UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Deep Learning Approaches for Cell Segmentation and Tracking in Time-Lapse Microscopy

A dissertation submitted in partial satisfaction of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

By

Abolfazl Zargari

June 2024

The Dissertation of Abolfazl Zargari
is approved:

Professor Ali Shariati, Chair

Professor Keith Corzine

Professor Zouheir Rezki

Professor Marcella Gomez

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Deep Learning Approaches for Cell Segmentation and Tracking in Time-Lapse Microscopy

Abolfazl Zargari

This dissertation addresses the challenges of cell segmentation and tracking in time-lapse microscopy using advanced deep-learning techniques. Analyzing microscopy images involves several challenges, including accurate segmentation and tracking of cells, handling the presence of artifacts and noise, and dealing with the proximity and overlapping of cells in densely populated images. Furthermore, there is a significant challenge in obtaining large annotated datasets necessary for training robust deep-learning models, as manually annotating microscopy images is time-consuming and labor-intensive. To overcome these issues, first, we developed DeepSea, a deep-learning model for efficient cell segmentation and tracking. DeepSea incorporates auxiliary models for cell edge detection, residual blocks for efficiency, and progressive learning techniques, achieving high segmentation accuracy and effective cell tracking. Additionally, we propose cGAN-Seg, a CycleGAN-based model that generates synthetic images to enhance the training process of cell segmentation models, incorporating style generation paths, linear attention mechanisms, differentiable image augmentation, and VGG perceptual loss. This significantly improves segmentation performance on limited datasets, with substantial improvements across various cell types and imaging modalities. A GAN-based super-resolution video generator is also introduced, generating annotated high-quality, realistic time-lapse microscopy videos, further addressing annotated dataset scarcity for live single-cell tracking models. Finally, we employed our quantitative single-cell image analysis pipeline to gain insights into cell size regulation and morphological diversity, as well as cell spatial and frequency feature distribution.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Professor Ali Shariati, for his unwavering support, guidance, and encouragement throughout my PhD journey. I am truly grateful for his mentorship and the opportunities he has provided me to grow as a researcher.

Finally, thank you to Najmeh for your love, being my best friend, and being my biggest supporter. And thank you to my family and friends for all your love and support.

During my PhD studies, I had the privilege of contributing to the following publications:

1. A. Zargari, G. A. Lodewijk, N. Mashhadi, N. Cook, C. W. Neudorf, K. Araghbidikashani, R. Hays, S. Kozuki, S. Rubio, E. Hrabeta-Robinson, A. Brooks, L. Hinck, and S. A. Shariati, "DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy," *Cell Rep. Methods*, vol. 3, no. 6, p. 100500, Jun. 2023, doi: 10.1016/j.crmeth.2023.100500.
2. A. Zargari, B. R. Topacio, N. Mashhadi, and S. A. Shariati, "Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks," *iScience*, vol. 27, no. 5, p. 109740, May 2024, doi: 10.1016/j.isci.2024.109740.
3. A. Zargari, N. Mashhadi, and S. A. Shariati, "Enhanced Cell Tracking Using A GAN-based Super-Resolution Video-to-Video Time-Lapse Microscopy Generative Model," *bioRxiv*, Jun. 2024, doi: 10.1101/2024.06.11.598572.
4. A. Zargari Khuzani, M. Heidari, and S. A. Shariati, "COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images," *Sci. Rep.*, vol. 11, p. 9887, May 2021, doi: 10.1038/s41598-021-88807-2.

Introduction

1.1 Background

Live single-cell imaging allows biologists to directly capture the dynamic of cellular processes with high temporal resolution. This powerful tool allows for the visualization of cell biological processes such as including cell division, growth, migration, and intracellular signaling, which are fundamental to understanding biological functions and disease mechanisms. However, accurate quantification of microscopy images in an automated manner remains challenging. New frontiers are emerging in the integration of artificial intelligence (AI) and machine learning (ML) with advanced microscopy techniques to automate the analysis of live-cell imaging experiments with the goal of providing novel insights into cellular behaviors.

1.2 Motivation

The motivation for this dissertation stems from the need to address several challenges associated with live-cell imaging, particularly in the context of cell segmentation, tracking, and analysis. Cell segmentation refers to the process of partitioning a microscopy image into segments that correspond to individual cells, which is crucial for subsequent analysis of cell morphology and behavior. Cell tracking, on the other hand, involves monitoring and recording the movement and lineage of individual cells over time in a series of time-lapse images, enabling the study of dynamic cellular processes. Traditional methods often struggle with the complexities of live-cell imaging data, such as variations in cell morphology, dynamic cellular behaviors, and the presence of noise and artifacts in microscopy images. The advent of AI and ML offers promising solutions to these challenges, enabling more accurate and efficient analysis of live-cell imaging data.

1.3 Objectives

The primary objective of this dissertation is to develop and evaluate advanced deep-learning tools for the segmentation, tracking, and quantitative analysis of single cells in live-cell imaging experiments. The specific aims include:

Developing DeepSea: A versatile and trainable deep learning model for single-cell segmentation and tracking in time-lapse microscopy images. This model aims to automate the analysis of cellular features and provide insights into cellular dynamics and heterogeneity.

Introducing cGAN-Seg: A CycleGAN-based segmentation model designed to enhance cell segmentation performance with limited annotated training data. This approach seeks to overcome the limitations of existing segmentation models by generating realistic and diverse synthetic microscopy images.

Creating tGAN: A GAN-based super-resolution video-to-video generative model for time-lapse microscopy. This model aims to generate high-quality synthetic annotated datasets, facilitating the training of deep learning models for various cell analysis tasks.

Quantitative Analysis: Conducting a quantitative analysis of single-cell images using computational tools and algorithms. This analysis aims to characterize the morphological and functional diversity of different cell types and understand the underlying biological mechanisms.

1.4 Structure of the Dissertation

This dissertation is structured into several chapters, each addressing a specific aspect of live-cell imaging and the application of AI and ML techniques:

Chapter 2: Live-Cell Imaging: Provides an overview of live-cell imaging, including its principles, techniques, applications, and challenges.

Chapter 3: Integration of AI and ML in Microscopy Image Analysis: Discusses the fundamentals of AI and ML, data preparation, machine learning techniques, and the innovations enabled by these technologies in microscopy.

Chapter 4: DeepSea: Introduces the DeepSea model for single-cell segmentation and tracking, detailing its design, implementation, and performance evaluation.

Chapter 5: cGAN-Seg: Presents the cGAN-Seg model for enhanced cell segmentation with limited training data, discussing its architecture, methods, and results.

Chapter 6: tGAN: Describes the tGAN model for generating super-resolution time-lapse microscopy videos, including its design, methods, and performance evaluation.

Chapter 7: Quantitative Analysis of Single-Cell Images: Explores the use of computational methods for the quantitative analysis of single-cell images, highlighting key findings and insights.

Chapter 8: Conclusion: Summarizes the key contributions of the dissertation and discusses future research directions.

1.5 Contributions

This dissertation makes several significant contributions to the field of live-cell imaging and microscopy image analysis:

DeepSea Model: Development of a deep learning model for efficient cell segmentation and tracking, demonstrating high accuracy and robustness in time-lapse microscopy images.

cGAN-Seg Model: Introduction of a CycleGAN-based approach for enhancing cell segmentation performance with limited annotated datasets, showing significant improvements over traditional methods.

tGAN Model: Creation of a GAN-based generative model for producing high-quality synthetic annotated time-lapse microscopy videos, facilitating more effective training of time-lapse microscopy deep learning models.

Quantitative Analysis Framework: Implementation of a framework for the quantitative analysis of single-cell images, providing valuable insights into cellular heterogeneity and dynamics.

1.6 Summary

The integration of AI and ML in live-cell imaging represents a transformative advancement in cellular biology, enabling more accurate and detailed analysis of cellular processes. This dissertation addresses key challenges in the field by developing innovative deep-learning models and techniques, paving the way for future research and applications in biomedical imaging. The following chapters will delve into the specifics of these models and their contributions to the advancement of live-cell imaging.

2 Live-Cell Imaging

2.1 Introduction

Direct observation of single cells through live imaging can reveal dynamics of fundamental cellular processes such as cell division, cellular migration, and growth in real time using sophisticated imaging technologies. Live cell imaging enables scientists to directly visualize cellular processes as they occur, which is crucial for understanding the complex dynamics and functions within living organisms [1, 2]. By using fluorescent markers and fluorescent microscopes, researchers can monitor specific proteins and organelles within cells, tracking their movement and interactions over time at single-cell resolution. This dynamic approach offers a significant advantage over other single-cell methods or static imaging methods, which only provide snapshots of cellular activity, limiting the understanding of cellular processes to discrete moments.

The roots of live-cell imaging trace back to the 17th century when Antonie van Leeuwenhoek used his handcrafted microscopes to reveal the "uncanny world of the very small," providing the first glimpses of living cells and microorganisms (Figure 2.1). However, the field truly advanced in the 1950s and 1960s with the development of phase-contrast microscopy by Frits Zernike and the subsequent invention of fluorescence microscopy [3]. These innovations allowed cells and cellular components to be seen with greater clarity and contrast, revolutionizing the study of live cells [4]. In the 1990s, the introduction of green fluorescent protein (GFP) as a biological marker was a landmark

event. It enabled the visualization of protein locations, movements, and interactions within live cells without disrupting their normal functions [5]. This discovery earned Osamu Shimomura, Martin Chalfie, and Roger Y. Tsien the Nobel Prize in Chemistry in 2008, highlighting its significance in biological research. Over the years, live-cell imaging has become indispensable in cellular biology. It has facilitated groundbreaking studies in cell division, migration, and signal transduction and contributed to medical advances in understanding diseases like cancer and neurodegenerative disorders [6]. The ability to observe cellular mechanisms in their natural state has not only provided insights into cellular function and behavior but also accelerated drug discovery and development processes by allowing real-time monitoring of cellular responses to new treatments [7]. Moreover, technological advancements in live-cell imaging techniques continue to expand its capabilities. Innovations such as super-resolution microscopy and automated image analysis using artificial intelligence are pushing the boundaries of what can be visualized and understood from living cells [8]. These advancements are enhancing the temporal and spatial resolution with which researchers can observe cellular processes, thus opening new avenues for discovery in cellular biology and beyond [9].

Figure 2.1: Dutch cloth merchant named Antonie van Leeuwenhoek. He created single-lens microscopes with very high magnification. Using these lenses, he was the first to see and describe tiny living things called 'animalcules.'

2.2 Principles of Live-cell Imaging

2.2.1 Basic Principles of Microscopy as Applied to Live-cell Imaging

Live-cell imaging utilizes the fundamental principles of microscopy to observe living cells in real-time while preserving their natural physiological processes. Key to this method is the use of optical systems that minimize phototoxicity as well as an environmental control system to keep cells alive during imaging. This involves the careful selection of objectives, light sources, and imaging modalities to ensure optimal resolution and minimal disruption to cell function. In addition, an environmental control chamber should be coupled to the microscope to regulate the temperature and CO2 of the cells.

Optical Systems: The optical system in live-cell imaging typically includes high numerical aperture objectives, which provide a larger collection efficiency for emitted light from the sample, enhancing image clarity and detail. Additionally, the use of sensitive cameras capable of capturing images under low-light conditions is crucial to reduce exposure times and minimize light-induced damage to the cells.

Light Sources: Choosing the appropriate light source is vital for live-cell imaging. LEDs and lasers are commonly employed for their precise control over intensity and wavelength, which helps in reducing phototoxic effects associated with prolonged exposure to high-energy light.

2.2.2 Contrast Methods Used in Live-cell Imaging

Contrast enhancement is critical for visualizing cellular structures and functions that are otherwise transparent. Live-cell imaging employs various contrast methods to aid in this visualization:

Phase Contrast Microscopy: This technique is widely used to enhance the visibility of transparent and colorless specimens, making it ideal for observing live cells without the

need for staining (Figure 2.2). Additionally, phase contrast imaging is non-toxic, ensuring that the cells remain viable and behave naturally during the experiments. Phase contrast microscopy converts the phase shifts in light passing through the specimen into brightness changes in the image, effectively highlighting cellular structures [11].

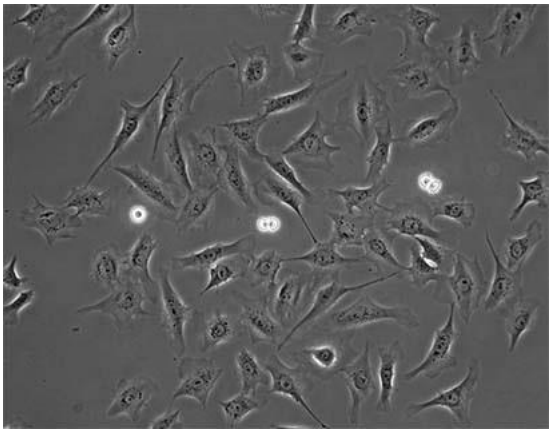

Figure 2.2: Phase contrast microscopy of Rat cells [12].

Fluorescence Microscopy: Fluorescence microscopy is another pivotal technique in live-cell imaging (Figure 2.3). It involves labeling cell components with fluorescent dyes or proteins and exciting them with specific wavelengths of light. The emitted light from the fluorophores is then captured to form an image, allowing for detection of signal from a specific protein or cellular organelle [13].

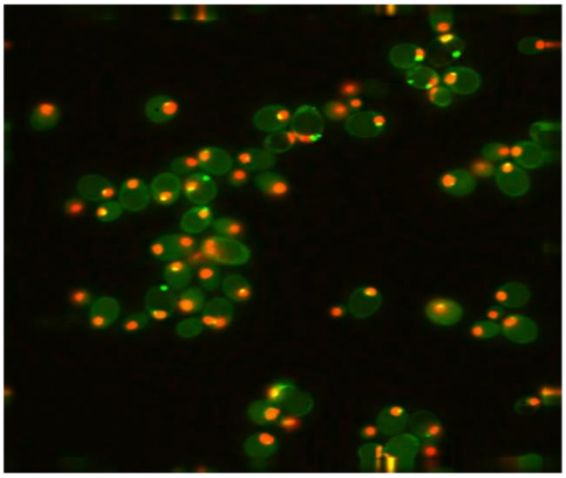

Figure 2.3: Fluorescence microscopy of budding yeast cells progressing through meiosis [14].

Differential Interference Contrast (DIC): DIC enhances contrast in unstained, transparent samples by utilizing differences in light refraction through the specimen (Figure 2.4). It produces a pseudo-three-dimensional effect that is useful for detailed morphological studies of living cells [15].

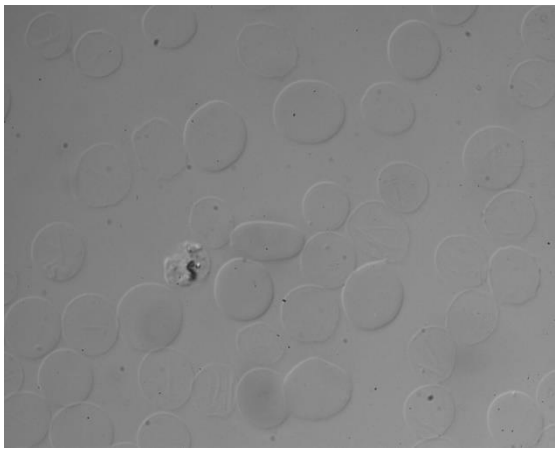

Figure 2.4: Differential interference contrast images of red blood cells [16].

2.2.3 Importance of Maintaining Cell Viability During Imaging

Maintaining cell viability during live-cell imaging is paramount for accurate interpretation of cellular dynamics and functions. Several strategies are implemented to ensure cells remain healthy and active throughout the imaging process [17]:

Temperature Control: Live-cell imaging systems are often equipped with temperature-controlled stages or enclosures that maintain a physiological temperature conducive to normal cell function.

Medium and Environment: The imaging medium should closely mimic the natural environment of the cells, including appropriate buffers, pH, and osmolarity. Additionally, CO_2 levels must be regulated when imaging cells that require a controlled atmospheric composition.

Minimizing Phototoxicity and Mechanical Stress: Reducing the intensity and duration of light exposure is crucial to prevent phototoxic effects that can alter cell behavior or induce cell death. Similarly, minimizing mechanical disturbances during setup and focusing is critical to avoid physical stress on the cells.

In conclusion, while setting up live-cell imaging experiments can be a complex process, it can be a highly informative technique that provides insights into the dynamic nature of cellular processes. By integrating advanced microscopy techniques with careful experimental preparation, researchers can observe and analyze cellular functions in unprecedented detail and with high temporal resolution l while ensuring the health and viability of the cells being studied.

2.3 Imaging techniques to study live cellular processes

There are several advanced microscopy methods that are pivotal for conducting live-cell imaging studies. Each technique offers unique advantages and is suitable for different types of biological inquiries. Some of the widely used methods are introduced as follows:

Fluorescence Microscopy: Fluorescence microscopy is one of the most common techniques used in live-cell imaging. It involves labeling cell components with fluorescent tags and exciting these tags with specific wavelengths of light. The emitted light is then captured to produce images that reveal the distribution and dynamics of cellular constituents. This method is invaluable for tracking cellular events over time, although it can be limited by photobleaching and phototoxicity, which may affect cell viability [18].

Confocal Microscopy: Confocal microscopy enhances the clarity of fluorescence imaging by using a spatial pinhole to block out-of-focus light in specimens that are thicker than a single optical section. This results in images with superior resolution and contrast. Confocal microscopy is particularly useful for detailed three-dimensional reconstructions of cellular structures, enabling researchers to precisely observe the interactions and functions of various cellular components within a thicker context [19].

Two-photon and Multi-photon Microscopy: Two-photon microscopy uses near-infrared excitation light, which allows deeper penetration into the specimen with less photodamage compared to traditional fluorescence microscopy. This is advantageous for imaging thicker tissues. Multi-photon microscopy extends this principle by using two or more photons of lower energy to excite the fluorophore simultaneously, further reducing phototoxicity and enabling detailed imaging of live animals over extended periods [20].

Total Internal Reflection Fluorescence Microscopy (TIRF): TIRF microscopy utilizes an evanescent wave that only illuminates a thin section of the specimen immediately adjacent to the glass-water interface, typically less than 100 nanometers thick. This technique is particularly suited for studying cell membrane dynamics, such as vesicle trafficking and receptor-ligand interactions at the cell surface [21].

Super-resolution Microscopy: Techniques such as Stimulated Emission Depletion (STED), Photoactivated Localization Microscopy (PALM), and Stochastic Optical Reconstruction Microscopy (STORM) break the diffraction limit of light to provide resolution at the nanoscale. These methods allow researchers to observe structures and events in cells in unprecedented detail, far beyond what traditional light microscopy can achieve [22].

Time-lapse Microscopy: Time-lapse microscopy involves capturing a sequence of images at specific intervals over time to track the dynamic processes within live cells. This method is crucial for understanding cellular processes such as mitosis, migration, and intracellular trafficking over extended periods [23].

High-content Screening (HCS): High-content screening combines automated microscopy with quantitative analysis to gather data from large sets of cellular images. This technique is highly effective for screening large libraries of compounds or genetic modifiers in drug discovery and genetic research. HCS allows for the simultaneous analysis of multiple cellular parameters, providing a comprehensive overview of cellular responses to experimental treatments [24].

Each of these techniques has been optimized to address specific scientific questions and experimental conditions, making them indispensable tools in the field of cell biology and medical research. Together, they provide a comprehensive suite of options for

researchers looking to explore the complexities of cellular function and disease pathology in a living context.

2.4 Image Processing and Analysis Software Tools

Quantitative analysis of live-cell imaging relies heavily on sophisticated image processing and analysis software. These tools are essential for transforming raw imaging data into meaningful quantitative information that can inform biological hypotheses. Key functionalities of these software packages include noise reduction, signal enhancement, object recognition, and tracking capabilities. Examples of widely used software in live-cell imaging include ImageJ/FIJI, which is favored for its plugin-rich, open-source platform; MATLAB and its Image Processing Toolbox for customizable processing workflows; and commercial software like MetaMorph and Volocity, which provide integrated solutions with user-friendly interfaces [25, 26].

Python has also become increasingly popular for image processing in live-cell imaging due to its extensive libraries, such as OpenCV, scikit-image, and NumPy, which facilitate powerful and flexible image analysis [27]. Each software choice has strengths specific to the types of analyses. For instance, ImageJ is particularly useful for general image processing and analysis, while MATLAB and Python excel in handling custom algorithm development for specific applications like fluorescence resonance energy transfer (FRET) or fluorescence recovery after photobleaching (FRAP) analyses [28]. Proper selection and application of these tools are crucial for effective data interpretation, requiring users to have both technical knowledge of the software and biological insights into the sample being studied [29].

2.5 Applications of Live-cell Imaging

Live imaging offers the possibility of recording cellular processes of single cells at high temporal resolution, which allows for the dissection of dynamics of heterogenous

biological processes that are often masked in bulk and static analysis. Here, I focus on a few applications of live cell imaging as pertains to my PhD dissertation:

2.5.1 Study of Cell Dynamics

Cell Division Cycle: live-cell imaging is instrumental in exploring the various phases of the cell cycle in a dynamic and uninterrupted manner. Fundamental to the continuity of life is the ability of individual cells to divide into two cells. This orchestrated process is the cell division cycle and has four distinct phases (Figure 2.5). Each phase of the cell cycle—G1, S, G2, and M—can be studied in detail [39]:

- G1 Phase (Gap 1): This phase involves cellular growth and the synthesis of mRNA and proteins necessary for DNA replication. Live-cell imaging allows for the observation of changes in cell size and the preparation for DNA synthesis.
- S Phase (Synthesis): DNA replication occurs during this phase, and using fluorescence microscopy, researchers can monitor the replication of DNA and the synthesis of additional chromosomal components. This phase is crucial for ensuring genetic material is accurately copied, and any errors are promptly repaired.
- G2 Phase (Gap 2): Following DNA synthesis, the cell prepares for mitosis. This phase can be monitored to study protein synthesis and the mechanisms that ensure the cell's readiness for mitosis, including the integrity of the DNA after replication.
- M Phase (Mitosis): This phase, which includes prophase, metaphase, anaphase, and telophase, can be intricately observed to track the organization and segregation of chromosomes. Live-cell imaging facilitates the study of mitotic spindle dynamics and the alignment of chromosomes, which are critical for ensuring accurate cell division.

Utilizing time-lapse microscopy, researchers can directly observe the chronological progression of cell division in real-time across different cell types and treatments. This method allowed for the direct observation and quantitative analysis of progression through different phases of the cell cycle. Changes in size and shape of the single cells can easily be revealed through label-free, non-toxic phase-contrast imaging [36]. In addition, key mitotic events, such as the formation of chromosomes and their alignment at the metaphase plate, can also be monitored by live single-cell imaging. This real-time data from live imaging experiments can also reveal the emergence of cancer cells in the dish, where the regulation of cell division is often compromised, leading to uncontrolled cell proliferation.

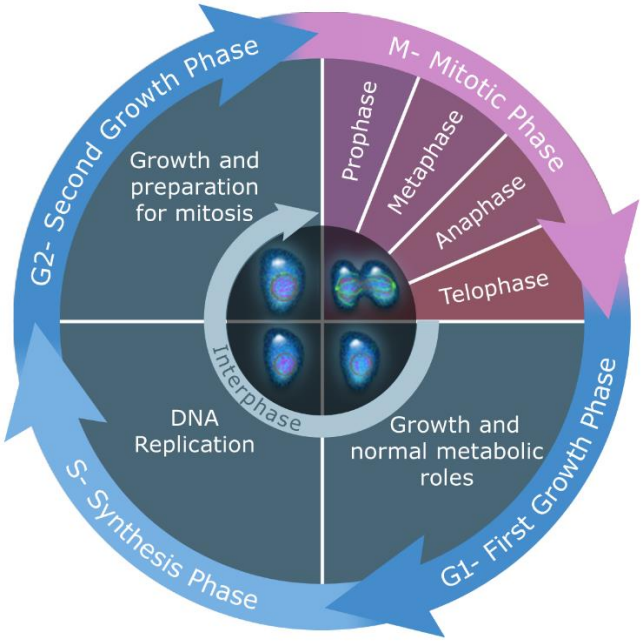

Figure 2.5: Cell Cycle Phases [40].

Cell Migration: Another pivotal application of live-cell imaging is in the study of cell migration, which is integral to processes such as wound healing, immune responses, and

the invasion and metastasis of cancer cells. Live imaging techniques enable researchers to track the movement of cells across a substrate in real-time, offering insights into the mechanics of cell motility and the roles of various signaling pathways that govern directional movement. Advanced imaging modalities, including fluorescence and confocal microscopy, are particularly useful for elucidating the functions of specific molecules in cell migration. These techniques allow for the visualization of intracellular components in high resolution, aiding in the identification of cytoskeletal elements and adhesion molecules that play crucial roles in the migration process.

By integrating live-cell imaging techniques to study these cellular dynamics, scientists have gained a deeper understanding of how cells use growth signals to control their division and how cells may form memories about past exposure to growth signals [38]. This method not only provides a visual and quantitative analysis of cellular processes under psychological conditions but also offers novel insights that lead to unregulated cell division in cancer cells.

2.5.2 Investigation of Intracellular Processes

Signal Transduction: Live-cell imaging is used to investigate the dynamics of signal transduction pathways within cells. This includes observing the dynamics of response to external signals and how different dynamics are converted into a cellular decision [41]. In these experiments, fluorescent reports are designed to change their level or localization in response to specific external signaling factors. Real-time imaging helps visualize the dynamics of protein interactions and movements across the cell membrane and within the cell, enhancing our understanding of cellular communication and function.

Organelle Trafficking: Another significant application is the study of organelle trafficking. Live-cell imaging tracks the movement and interaction of organelles like mitochondria, lysosomes, and the Golgi apparatus within cells. Such imaging is crucial for understanding

organelle function, dynamics, and their role in diseases such as neurodegenerative disorders, where organelle transport is often impaired [42].

2.5.3 Drug Development and Cellular Response Analysis

Screening and Development: In the pharmaceutical industry, live-cell imaging is employed for drug screening and development. This method allows for the direct observation of how drugs affect cells, providing essential data on drug efficacy and toxicity before clinical trials. For example, researchers can monitor how cancer cells respond to new chemotherapy agents or assess the side effects of drug candidates in real time [43].

Mechanism of Action: Understanding the mechanism of action of drugs is facilitated by live-cell imaging. This application is critical in determining how drugs interact with cells and their components and what changes occur at the molecular level. This information can lead to the optimization of drug dosages and schedules, potentially reducing side effects and enhancing therapeutic efficacy [44].

Disease Modelling and Pathology: Live-cell imaging is invaluable in disease modeling. By observing how cells react to pathogenic infection or genetic mutations, researchers can elucidate disease mechanisms and identify potential therapeutic targets. This real-time data is critical for the development of treatments for complex diseases such as Alzheimer's and other chronic conditions [45].

In conclusion, live-cell imaging, with its ability to track cellular and subcellular activities in real time, continues to be a cornerstone in the field of life sciences. As imaging technologies evolve, they will further enhance our understanding of cell biology, aid in the development of new drugs, and offer profound insights into cellular responses to therapies. This technique not only allows us to visualize the unseen dynamics of life but also holds

the promise of unlocking new therapies and treatments that can manage or cure diseases at a cellular level.

2.6 Challenges and Limitations of Live-cell Imaging

2.6.1 Technical and Practical Challenges in Live-cell Imaging

Live-cell imaging is a powerful technique used to observe the dynamic processes of living cells in real time. However, this technique poses various technical and practical challenges. The primary technical challenge is maintaining cell viability and normal physiology during prolonged imaging sessions. Live cells are highly sensitive to their environment; factors such as temperature, pH, and ion concentrations must be strictly controlled to prevent cell damage or death [46]. Additionally, the light exposure required for imaging, particularly in fluorescence microscopy, can lead to phototoxicity and photobleaching. Phototoxicity can alter cell behavior or even kill cells, while photobleaching reduces the fluorescence signal over time, complicating data acquisition [47].

Another practical challenge is the complexity of the imaging setup. Advanced live-cell imaging systems, such as confocal microscopes or high-throughput imaging platforms, require significant expertise and calibration to operate effectively. Ensuring consistent focus over time and across different focal planes (z-stacking) also demands sophisticated hardware and software solutions, such as adaptive focus control [48].

2.6.2 Limitations in Spatial and Temporal Resolution

The spatial and temporal resolution of live-cell imaging is another critical limitation. Spatial resolution is restricted by the diffraction limit of light, which for conventional optical microscopy is about 200 nanometers. This makes it difficult to visualize structures smaller than this limit, such as individual molecules or smaller organelles. Techniques like super-resolution microscopy have been developed to overcome this limit, but they often require more intense light or more complex procedures, which can exacerbate phototoxicity [50].

Temporal resolution is limited by the speed at which images can be acquired without compromising image quality. Faster imaging is necessary to capture rapid cellular processes, but increased speed can reduce image resolution and increase noise, making it harder to discern fine details. Balancing these factors is a constant challenge in designing live-cell imaging experiments [51].

2.7 Future Perspectives

2.7.1 Predicted Developments in Imaging Technologies

The evolution of live-cell imaging technologies continues to advance at an unprecedented pace, driven by the need to visualize and understand cellular processes in real time and in their natural environment. Innovations in high-resolution microscopy, such as super-resolution and light-sheet microscopy, are anticipated to grow further. These technologies aim to reduce phototoxicity, enabling longer observation periods while maintaining cellular viability [51].

Another significant development is the improvement in time-lapse imaging. This technique is likely to see advances in temporal resolution which will allow researchers to capture faster biological processes accurately. Coupled with advancements in fluorescent protein engineering and the emergence of novel biosensors, these developments are expected to enhance our ability to monitor dynamic cellular events with minimal intrusion [52].

Quantitative phase imaging (QPI) technologies, which provide high-contrast, label-free imaging, are set to become more mainstream. These methods will benefit from computational advancements, potentially allowing for the non-invasive analysis of cellular components and their physical properties over time, contributing to a deeper understanding of cell biology without the need for dyes or labels [53].

2.7.2 Potential New Applications in Biomedical Research

As imaging technologies evolve, so too will their applications in biomedical research. We are likely to see increased use of live-cell imaging in drug discovery, where understanding the dynamics of drug interactions at the cellular level can lead to the development of more effective therapeutics [54]. Additionally, the integration of live-cell imaging with organ-on-a-chip models could revolutionize toxicity testing and disease modeling by providing more accurate simulations of human physiological responses [55].

Live-cell imaging is also poised to expand its role in immunology, particularly in studying the interactions between immune cells and pathogens in real time. This could significantly enhance our understanding of infection mechanisms and immune response, leading to better vaccines and treatments [56].

Furthermore, the adaptation of live-cell imaging for use in personalized medicine is on the horizon. By observing how individual cells from a patient respond to therapeutic agents, personalized treatment regimens can be developed and optimized for efficacy and minimal side effects.

2.7.3 Integration of Artificial Intelligence and Machine Learning in Image Analysis

Artificial Intelligence (AI) and Machine Learning (ML) are set to revolutionize image analysis in live-cell imaging by improving the speed, accuracy, and efficiency of data processing. Deep learning algorithms, in particular, are making it possible to automate complex image analysis tasks that were previously labor-intensive and prone to human error [57].

One of the forefront applications of AI in live-cell imaging is in feature detection and classification. AI models can be trained to recognize patterns and anomalies in cell behaviors, facilitating faster and more accurate diagnoses of diseases such as cancer [58].

Moreover, these technologies are enhancing the quantification of cellular components, enabling researchers to gather precise measurements of cell morphology, motility, and interaction dynamics [59].

AI and ML are also integral in managing and interpreting the vast amounts of data generated by modern imaging techniques. By leveraging these tools, researchers can uncover subtle biological processes and interactions at a scale and with a level of detail that was previously unattainable. In addition, predictive modeling using AI can forecast outcomes of cellular processes under various experimental conditions, thereby not only saving valuable research time but also minimizing resource expenditure.

In conclusion, the integration of emerging imaging technologies with AI and ML is not only expanding the capabilities of live-cell imaging but is also paving the way for groundbreaking discoveries in biomedical research. The future of live-cell imaging promises to bring even more sophisticated tools and methodologies that will continue to push the boundaries of what we can observe and understand about living cells.

2.8 Conclusion

Live-cell imaging has revolutionized cellular and molecular biology by providing real-time visualization of dynamic cellular processes. Utilizing high-resolution microscopy and fluorescent markers, it has advanced our understanding of cell division, migration, signal transduction, and organelle interactions, significantly impacting medical research and drug development.

This technology allows for the observation of cellular processes in their natural state, enhancing our understanding of the mechanisms underlying health and disease. It has been instrumental in discoveries across various fields, such as neuroscience and oncology, by enabling researchers to monitor processes like cancer cell migration and

response to therapies in real time, thus improving experimental accuracy and manipulation of biological systems.

Future innovations in live-cell imaging, including improved super-resolution techniques, adaptive optics, and AI integration, promise to refine observation capabilities without compromising cell health. Combining live-cell imaging with technologies like CRISPR-Cas9, organ-on-a-chip systems, and single-cell sequencing will further dissect cellular behavior, enhancing drug testing, personalized medicine, and disease understanding. Live-cell imaging remains a cornerstone of biological sciences, driving new discoveries and medical innovations.

3 Integration of Artificial Intelligence and Machine Learning in Microscopy Image Analysis

3.1 Introduction

The advent of artificial intelligence (AI) and machine learning (ML) in recent years has transformed many scientific domains, notably within biological research, where microscopy serves as an essential tool to visualize cellular and molecular structures. This integration offers unprecedented capabilities for automated image analysis, ultimately accelerating discoveries and innovations [61, 62].

Artificial Intelligence (AI) and Machine Learning (ML) are branches of computer science that emphasize the creation and implementation of algorithms that can learn from and make predictions or decisions based on data (Figure 3.1). AI incorporates a broader concept of machines being able to carry out tasks in a way that we would consider "smart", while ML is a specific subset of AI that trains a machine how to learn from data [63].

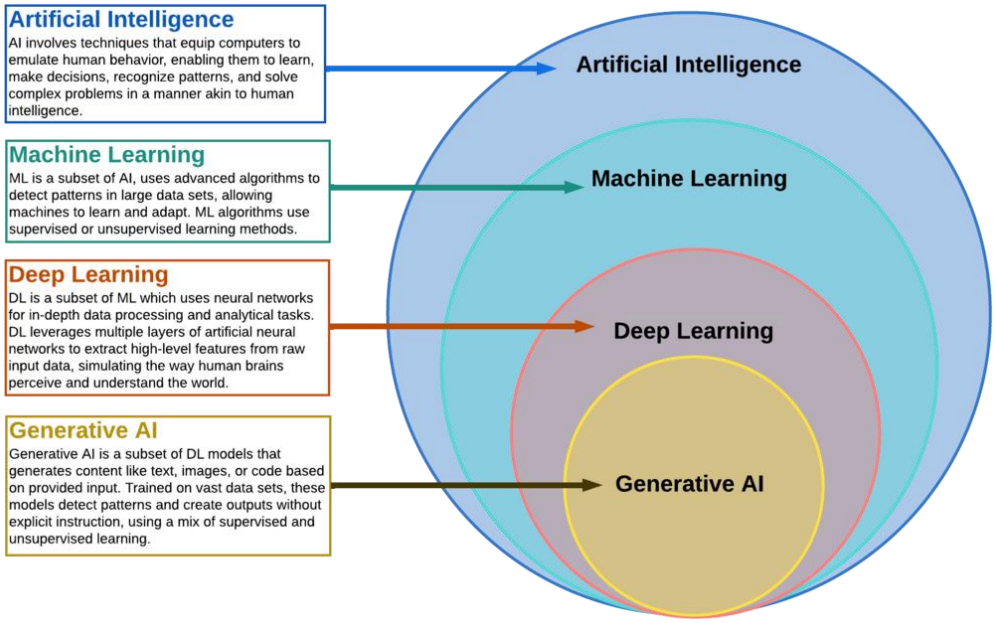

Figure 3.1: Unraveling AI Complexity: A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI [64].

In the context of microscopy image analysis, AI and ML have become transformative tools. Machine learning, in particular, has been crucial in developing methods that can automatically recognize complex patterns in image data—something that is incredibly challenging for traditional algorithms. These methods include supervised learning techniques, where the system learns from a dataset containing labeled images to create a predictive model, and unsupervised learning techniques, which identify patterns and structures in unlabeled data [65, 66].

AI and ML not only improve the accuracy and efficiency of image analysis but also enhance the ability to handle large datasets—common in modern high-throughput and high-content screening methods. By automating the process of feature detection, segmentation, and classification, AI-driven tools can analyze vast quantities of data with a

consistency that is not feasible for human observers. Moreover, these tools can adapt and improve over time as they are exposed to more data [67].

Furthermore, deep learning, a particular type of machine learning characterized by deep neural networks, has shown significant promise in the field of image analysis. These networks excel in tasks such as feature identification and image classification, outperforming traditional machine-learning models in many cases. Their ability to learn progressively more complex features at various levels of abstraction allows them to make sense of data with high variability and detail, such as those encountered in microscopy images of biological tissues [68].

This integration of AI and ML into microscopy image analysis not only streamlines workflows but also opens up new avenues for biological discovery, facilitating detailed phenotypic profiling, complex morphological categorization, and dynamic bioprocess analysis. As these technologies continue to evolve, they promise to further revolutionize our understanding of biological systems at the microscopic level [69].

3.2 Fundamentals of AI and ML in Image Analysis

3.2.1 Basic Principles of Machine Learning and Artificial Intelligence

Machine Learning relies on statistical methods to enable machines to improve at tasks with experience. The basic principle is to build models from input data that can predict outcomes or categorize data into different labels. These models are trained using a large set of data known as training data, which helps the algorithm adjust and improve its accuracy over time through a process called learning.

Artificial Intelligence extends beyond learning from data; it also involves reasoning, problem-solving, and decision-making capabilities that mimic human intelligence. This includes not only ML but also logic, planning, and symbolic reasoning.

3.2.2 Overview of Common Algorithms Used in Image Analysis

In the context of microscopy image analysis, several ML algorithms are particularly prevalent:

Convolutional Neural Networks (CNNs): These are deep learning algorithms that are particularly powerful for analyzing visual imagery. CNNs automatically detect important features without any human supervision, using layers of processing units for feature extraction and output generation. This capability makes them exceptionally good for image classification, object detection, and more complex tasks like scene recognition [70, 71].

Decision Trees: These models use a tree-like model of decisions and their possible consequences. It's a type of supervised learning algorithm that is used for classification and regression tasks. In image analysis, decision trees can help in segmenting the image into different parts based on the features [72].

Random Forests: An ensemble of decision trees typically used to improve the robustness and accuracy of decision trees. Random forests combine multiple decision trees to produce a more accurate and stable prediction [73].

Support Vector Machines (SVM): SVMs are another supervised learning method used for classification or regression problems. In image processing, SVMs are used for classification tasks, including image categorization and object recognition [74].

3.2.3 Differences Between Traditional Image Analysis and ML-driven Methods

Traditional image analysis often involves manual feature selection and rule-based methods. For instance, in microscopy, traditional methods might require setting thresholds for color intensity, size filtering, and shape parameters to identify objects of interest.

ML-driven methods, on the other hand, leverage algorithms that can learn these features and rules from the data itself. This not only reduces the need for manual intervention but often results in more robust, scalable, and flexible analysis systems. For example, a CNN trained on microscopy images can learn to identify patterns that are far too complex for traditional methods to detect, such as subtle differences in cell morphology.

Scalability and Adaptability: ML methods can easily scale with the addition of new data, and they adapt better to new, unforeseen scenarios once they are retrained with updated data sets. This contrasts with traditional methods, which may require manual re-tuning or redesign when new types of images are analyzed.

Accuracy and Efficiency: ML algorithms often surpass traditional methods in terms of accuracy, especially in complex image datasets where the manual feature design might miss nuanced patterns that a machine can learn to recognize.

Overall, integrating AI and ML in microscopy image analysis not only enhances the efficiency and accuracy of analyses but also opens up new avenues for scientific discovery that were previously unattainable with traditional methods. AI and ML-based solutions are rapidly replacing rule-based tools in microscopy image analysis, fostering advancements in both diagnostic and research methodologies.

3.3 Data Preparation and Preprocessing

Data preparation and preprocessing are critical steps in the development of ML models, particularly in the field of microscopy image analysis. These steps significantly affect the performance, accuracy, and reliability of the ML algorithms employed. Below, we expand on the importance of data quality and quantity, various techniques for image data augmentation and normalization, and methods for labeling and annotating microscopy images.

3.3.1 Importance of Data Quality and Quantity in Machine Learning

Data Quality: High-quality data is paramount in machine learning. For microscopy images, this means the images must be in focus, high-resolution, and free from artifacts that could mislead the learning algorithm. Factors such as focus, uniform lighting, and staining consistency are crucial. Noise reduction and artifact removal are often necessary to improve data quality. Algorithms need accurate and representative data to learn effectively; otherwise, they may produce biased or generalized models that fail to perform in practical applications.

Data Quantity: Alongside quality, the quantity of data is equally crucial. A larger dataset provides a more comprehensive basis for training models, enabling them to learn a wide variety of features and patterns. However, in specialized fields like microscopy, obtaining large datasets can be challenging due to the labor-intensive nature of data collection. Techniques such as data augmentation can help overcome these limitations by artificially expanding the dataset.

3.3.2 Techniques for Image Data Augmentation and Normalization

Data Augmentation: This technique involves generating new training samples from existing data by applying random but realistic transformations to input images. For microscopy images, common augmentation techniques include rotation, scaling, flipping, and the use of elastic deformations. These manipulations help the model generalize better to new, unseen images by simulating different viewing conditions and orientations commonly encountered in real-world scenarios.

Normalization: Normalization is another crucial preprocessing step designed to standardize the intensity levels across multiple images. This process is important because variations in lighting conditions, staining intensity, and camera settings can affect the visual appearance of microscopy images. Techniques such as min-max normalization, Z-score

standardization, or histogram equalization are commonly used to ensure consistent image quality and help ML models focus on relevant features rather than being misled by intensity variations.

3.3.3 Methods for Labeling and Annotating Microscopy Images

Manual Labeling: The most reliable method for annotating microscopy images involves expert biologists or trained annotators who manually label the images. This process, while time-consuming and expensive, provides the highest quality data, especially for complex structures that require a nuanced understanding of the subject matter.

Semi-Automated Annotation: To alleviate the burden of manual labeling, semi-automated tools that combine human expertise and algorithmic assistance are used. These tools might include interactive segmentation techniques, where the annotator outlines rough boundaries and the software refines them, or classification assistants that suggest potential labels based on preliminary ML models.

Automated Annotation: Fully automated annotation involves using pre-trained models to label new datasets. These models are trained on previously annotated datasets and can rapidly annotate large volumes of data. However, their accuracy depends heavily on the quality and relevance of the training data used. Continuous validation by human experts is necessary to ensure the reliability of automated annotations.

In summary, the effectiveness of ML models in microscopy image analysis heavily relies on rigorous data preparation and preprocessing steps. By ensuring high-quality and abundant data through proper preprocessing techniques and by employing meticulous labeling and annotating methods, researchers can enhance the performance and applicability of their ML models to real-world microscopy challenges. This foundation is

critical not only for model development but also for advancing the field of digital pathology, developmental biology, and other areas reliant on microscopic imaging.

3.4 Machine Learning Techniques Applied to Microscopy

3.4.1 Supervised Learning Models for Classification and Regression Tasks

Supervised learning involves training a model on a labeled dataset, where each input image is associated with an output label (classification) or a continuous output (regression). In microscopy, these models are pivotal for tasks such as identifying cell types, diagnosing diseases from histological images, and quantifying biological structures [74, 75].

Classification Tasks: Deep learning models, particularly convolutional neural networks (CNNs), have revolutionized the field of image classification. In microscopy, CNNs can differentiate between various cell types or identify pathological changes in tissues. For example, researchers have developed models that accurately classify cancerous cells in histopathology images, aiding in early diagnosis and treatment planning [76, 77].

Regression Tasks: Regression models in microscopy are used for quantitative tasks, such as estimating the concentration of a substance in a sample or measuring cell features (e.g., size, shape). These models can also be used for predictive modeling, such as predicting the potential growth rate of cultured cells based on initial imaging [78].

3.4.2 Unsupervised Learning Models for Clustering and Dimensionality Reduction

Unsupervised learning does not require labeled data, making it suitable for exploratory analysis or situations where manual labeling is impractical.

Clustering: Clustering algorithms, such as k-means or hierarchical clustering, are used to group similar images or features without prior knowledge of the group labels. This

technique can be instrumental in discovering new biological phenomena or in segregating cell populations in flow cytometry data without predefined markers [80, 81].

Dimensionality Reduction: High-dimensional data typical in microscopy can be challenging to analyze and visualize. Dimensionality reduction techniques like Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) help in reducing the number of random variables under consideration, revealing the intrinsic patterns in the data. For instance, PCA can be used to reduce the dimensionality of image datasets before classification, while t-SNE is often employed to visualize complex data clusters in a two-dimensional space [82, 83].

3.4.3 Reinforcement Learning in Adaptive Imaging Techniques

Reinforcement learning (RL) involves training models to make a sequence of decisions. In microscopy, RL can be applied to develop adaptive imaging systems that optimize their parameters in real time for improved image quality or faster data acquisition [84, 85].

Adaptive Imaging Techniques: Autofocus Systems: RL can be utilized to train autofocus systems in digital microscopy, allowing the microscope to adaptively adjust focus across different sample sections, enhancing image clarity and reducing manual adjustments [86].

Dynamic Adjustment of Imaging Parameters: RL algorithms can dynamically adjust exposure time, light intensity, and other imaging parameters based on the observed sample characteristics to optimize image quality [87].

Active Learning for Sample Selection: RL can be employed in systems where the algorithm actively selects which samples to image next based on the most informative samples, thereby improving the efficiency of large-scale experiments [88].

These machine learning techniques offer significant advantages in microscopy, enabling more efficient, accurate, and detailed image analysis. Their integration into microscopy not only enhances existing methodologies but also opens up new avenues for scientific discovery and technological innovation in the field [89].

3.5 Enhancements and Innovations Enabled by AI and ML models in Microscopy

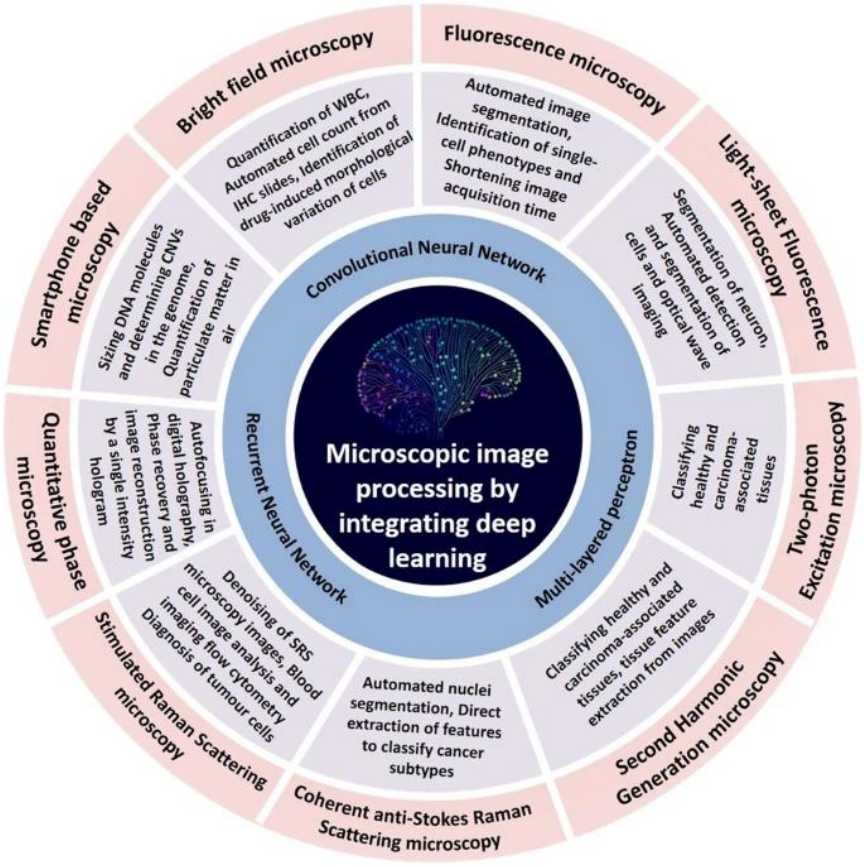

Figure 3.2: Deep learning-based image processing in optical microscopy [90].

3.5.1 Improvement in Image Resolution and Quality (Super-Resolution Microscopy)

Artificial Intelligence (AI) and Machine Learning (ML) have dramatically improved the resolution and quality of images captured by microscopes, particularly through super-resolution microscopy techniques. Super-resolution microscopy, which includes methods like STED, SIM, and PALM, traditionally faces limitations such as photobleaching and photon budget issues. AI and ML algorithms help overcome these by reconstructing high-resolution images from low-resolution data, thus enhancing the clarity and details visible in the images. Techniques such as deep learning have been used to predict high-resolution details from available data, reducing noise and increasing resolution beyond the diffraction limit of light. This has not only improved image quality but also allowed for the observation and analysis of biological processes at the molecular level, which were previously invisible with standard resolution methods [95, 96].

3.5.2 Automation of Repetitive Tasks

AI and ML excel in automating repetitive and labor-intensive tasks in microscopy image analysis, such as cell counting, size measurements, and morphological assessments. These tasks, which are tedious and error-prone when performed manually, can be handled efficiently and with greater accuracy by trained algorithms. Automation speeds up these processes significantly and allows researchers to focus on more complex analysis tasks. For instance, convolutional neural networks (CNNs) can be trained on a set of images to automatically recognize and count cells or measure their dimensions, even in cluttered or complex images where traditional methods fail. This automation not only saves valuable time but also reduces human error and enhances the reproducibility of experiments [97-99].

3.5.3 Enhanced Detection of Subtle or Complex Patterns Not Discernible by Human Observers

AI and ML are particularly adept at identifying patterns and anomalies in data that might be missed by human observers. In the context of microscopy, this capability allows for the detection of subtle or complex biological structures and phenomena. Advanced algorithms can analyze vast amounts of data to find patterns indicating disease, such as cancerous cells, or monitor the progression of a disease over time. Furthermore, unsupervised learning techniques can discover new patterns without prior knowledge or intervention, opening new avenues for scientific discovery. These capabilities extend beyond pattern recognition, as machine learning models can also predict outcomes based on data trends, thus providing valuable insights that can guide future research and clinical practices.

In summary, the integration of AI and ML into microscopy image analysis not only enhances the capabilities of traditional microscopy techniques but also transforms the scale and scope of what can be achieved. These advancements lead to more detailed and accurate scientific inquiry, better disease diagnosis, and potentially groundbreaking discoveries in biological research. By reducing the human workload and improving the precision of data analysis, AI and ML pave the way for a new era in microscopy.

3.5.4 Enhanced Health and Disease Diagnostics

AI and ML have revolutionized the field of health diagnostics through advanced microscopy techniques, particularly in the detection and analysis of diseases such as cancer and various pathogens.

Cancer Detection: Machine learning algorithms are increasingly used to automate the detection of cancer cells in histopathology images. Convolutional neural networks (CNNs), for example, have demonstrated high accuracy in distinguishing between benign and

malignant cells, significantly speeding up diagnostics and reducing human error. Studies have shown that AI can even identify cancer subtypes and predict patient outcomes by analyzing patterns invisible to the human eye [100, 101]. For instance, deep learning models have been effective in identifying clinically actionable genetic alterations in various cancers and predicting responses to treatments [102].

Pathogen Identification: In microbiology, AI techniques help in the rapid identification of bacteria, viruses, and parasites from microscopic images. Automated systems using image processing and machine learning can detect subtle morphological features that differentiate pathogens, facilitating faster and more accurate diagnoses than conventional methods. These advanced techniques are crucial for timely and effective treatment of infectious diseases, reducing the burden on healthcare systems [103].

3.5.5 Enhanced Drug Discovery and Development

The application of AI in microscopy extends significantly into the area of drug discovery, aiding in both the design and development of new pharmaceuticals.

High-Content Screening (HCS): High-content screening involves the automated analysis of large sets of cellular images using microscopy to assess the effects of thousands of potential drugs. AI algorithms optimize this process by analyzing image data to identify cellular changes specific to drug interactions, thereby predicting the efficacy and toxicity of pharmaceutical compounds. This method not only enhances the speed of analysis but also improves the predictive accuracy of drug responses [104-106].

3.6 Cell Segmentation

Cell segmentation and tracking are pivotal techniques in biological research, enabling scientists to observe and quantify the behavior of cells in dynamic and static states (Figure 3.3). These methodologies provide critical insights into cell morphology,

migration, division, and death, which are essential for understanding various physiological and pathological processes. Applications span numerous fields, including developmental biology, cancer research, neurology, and immunology, facilitating studies on cell development, disease progression, and treatment responses [107, 108].

The capability to isolate individual cells in an image (cell segmentation) and follow their trajectory over time (cell tracking) allows researchers to generate quantitative data on cellular dynamics and interactions. This not only aids in the visualization of cell behavior in their native environments but also supports high-throughput data analysis, which is crucial for statistical validation in experimental outcomes. As such, these techniques are instrumental in advancing personalized medicine, drug discovery, and the broader field of cell biology by providing a deeper understanding of cellular function and interaction [109, 110].

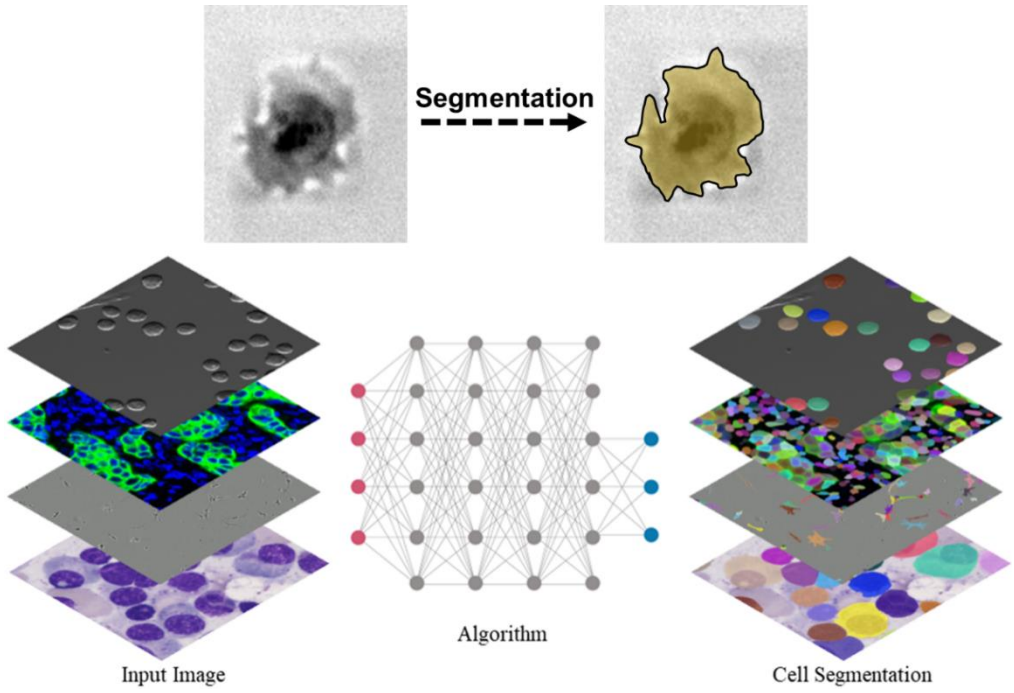

Figure 3.3: Cell segmentation algorithm [111]

3.6.1 Principles of Cell Segmentation

Cell segmentation is a fundamental process in digital image analysis where the goal is to partition a digital image into multiple segments to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. In the context of cellular biology, cell segmentation aims to accurately identify and outline individual cells within a heterogeneous image (Figure 3.3). This is crucial for quantitative analysis of cell morphology, tracking cell movements over time, and understanding cellular dynamics in different environments.

The primary goals of cell segmentation include:

Enhancing Image Interpretability: Making complex cellular images more understandable by isolating individual cells or groups of cells.

Quantitative Analysis: Facilitating the measurement of various cellular properties such as size, shape, texture, and intensity.

Automation: Reducing manual annotation efforts, thus speeding up the data processing pipeline in large-scale studies.

Accuracy and Precision: Ensuring high accuracy and precision in the identification and outlining of cells, which is critical for subsequent analysis like phenotyping or drug testing.

3.6.2 Overview of Common Techniques Used for Segmentation

Thresholding Methods: Thresholding is one of the simplest yet most effective approaches to segmenting images [112]. It involves converting a grayscale image into a binary image where the pixels are marked as either foreground (cell) or background. This is done by selecting a threshold value, and then all pixels above this threshold are classified as one part of the image, while those below are classified as another. There are various methods to determine the optimal threshold, including:

1- Global Thresholding: Suitable for images with high contrast between the foreground and background, where a single threshold is used across the entire image.
2- Adaptive Thresholding: Used for images with varying lighting conditions across different areas; it calculates thresholds for smaller regions, thus providing flexibility.
3- Otsu's Method: An automatic thresholding technique that determines the threshold by minimizing the intra-class variance of the black and white pixels.

Edge Detection Techniques: Edge detection techniques focus on identifying the boundaries of cells by detecting discontinuities in brightness [113]. Techniques include:

1- Sobel Operator: Used to detect edges based on the gradient of image intensity at each pixel.
2- Canny Edge Detector: A multi-stage algorithm that detects a wide range of edges in images.
3- Laplacian of Gaussian: This method applies a Gaussian blur to smooth the image before calculating the Laplacian, helping to reduce the noise and improve edge detection.

Region-Based Segmentation: Region-based segmentation techniques are aimed at finding regions within an image that are similar according to a set of predefined criteria [114]. Techniques include:

1- Region Growing: This method starts with a seed point and grows the region by appending to the region those neighboring pixels that have similar properties.
2- Watershed Algorithm: Often used for separating touch objects in an image, this technique is based on visualizing the image in three dimensions where brightness levels correspond to height.

Machine Learning Approaches: Machine learning approaches provide powerful tools for cell segmentation, especially in complex images where traditional methods fail [115, 116].

1- Supervised Learning: These methods require a labeled dataset and include techniques like support vector machines (SVM) and convolutional neural networks (CNN). They are trained to recognize cells based on features extracted from labeled examples.
2- Unsupervised Learning: These methods do not require labeled data and include algorithms like k-means clustering and hierarchical clustering that infer the structure of cell images based on the data provided.

Each of these techniques offers different advantages and is chosen based on the specific requirements of the image data and the analysis objectives. Combining multiple approaches can often provide better segmentation results, especially in complex imaging conditions or in the presence of heterogeneous cell populations.

3.6.3 Methods to Improve Segmentation Accuracy

Data Augmentation: By artificially increasing the diversity of training data through transformations like rotation, scaling, and flipping, models can generalize better to new images [117].

Advanced Post-processing: Techniques such as mathematical morphology can refine the segmentation results by removing noise and filling gaps in detected cell boundaries [118].

Ensemble Techniques: Combining the predictions from multiple models can reduce errors and improve the robustness of the segmentation [119].

Transfer Learning: Using a pre-trained CNN model and fine-tuning it on cell images can leverage learned features from extensive datasets, typically leading to better performance than training a model from scratch [120].

Implementing these computational techniques and evaluation strategies ensures the effective segmentation of cells, thereby facilitating accurate biological interpretations and findings.

3.7 Cell Tracking

3.7.1 Definition and Objectives of Cell Tracking

Cell tracking is the process of monitoring the movements and behavior of cells across a series of time-lapse microscopy images or video sequences (Figure 3.4) [121,

122]. The primary objective of cell tracking is to obtain quantitative data that can be analyzed to understand cellular dynamics, such as migration patterns, proliferation rates, and interactions with the microenvironment. This technique is critical in areas such as developmental biology, cancer research, and regenerative medicine, where understanding the mechanics of cell behavior can provide insights into complex biological processes.

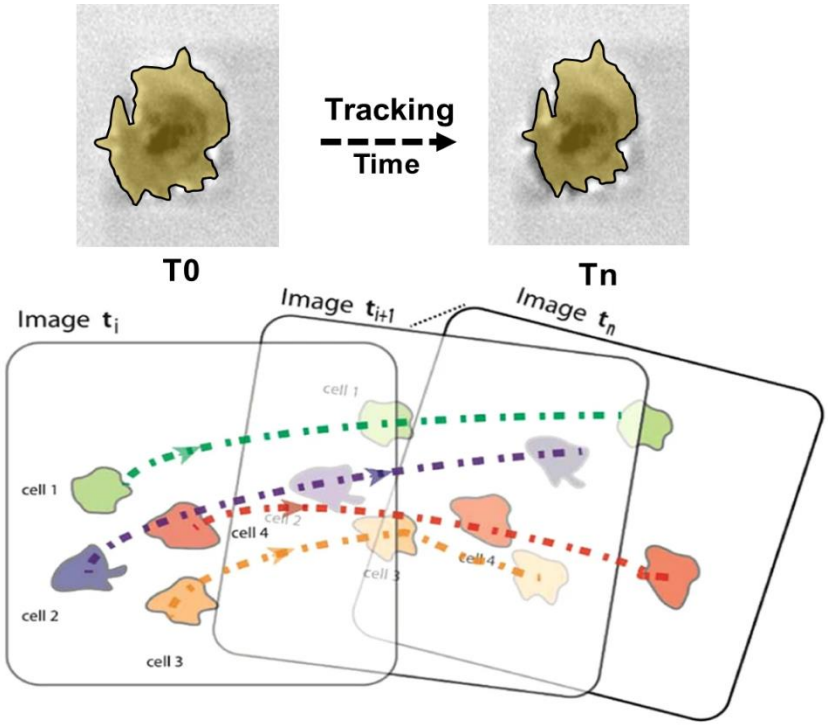

Figure 3.4: Cell tracking process [123].

Key objectives of cell tracking include:

Quantitative Analysis: Provide detailed, quantitative descriptions of cell movement, division, and death.

Behavioral Insight: Understand cellular responses under normal and experimental conditions, such as drug treatment or genetic modification.

Dynamic Modeling: Facilitate the development of models that predict cellular behavior under various physiological and pathological conditions.

3.7.2 Description of Challenges in Tracking Individual Cells Over Time

Tracking cells over time presents multiple technical challenges that can complicate the analysis:

Cell Division: The process of cell division creates a challenge in maintaining the identity of individual cells. Identifying which cells are the progeny of the division and tracking their subsequent paths is non-trivial.

Varying Intensities: Changes in cell appearance due to factors such as variable staining, photobleaching, or intrinsic changes in cell morphology can make it difficult to consistently identify the same cell over time.

Motion Blur: Rapid movement of cells, especially in vivo environments, can lead to motion blur in images, reducing the accuracy of tracking algorithms.

Cell Density: High cell density in cultures or tissues can lead to occlusions where cells overlap or touch, making it hard to differentiate and track individual cells.

3.7.3 Common Methodologies for Tracking

Single-Cell Tracking: Single-cell tracking focuses on following the path of individual cells. This technique is particularly useful in studying cell dynamics in a controlled environment where individual cell behavior plays a critical role in the overall outcome of the experiment. Methods often involve:

1- Manual Tracking: Direct human annotation of cell position in each frame, which is time-consuming but can provide highly accurate data for small datasets.

2- Automated Algorithms: Utilize edge detection, thresholding, and region-based methods to identify and track individual cells automatically through image sequences.

Multi-Cell Tracking: Multi-cell tracking deals with the simultaneous tracking of multiple cells within a collective group or tissue. This approach is essential for understanding the interactions within a cell population, such as in tumor growth or tissue development studies. Techniques include:

1- Graph-Based Methods: These create a graph where nodes represent detected cells in each frame and edges represent possible trajectories, optimized globally using various algorithms.

2- Model-Based Tracking: Involves using statistical models to predict cell positions and optimize tracking across frames based on these predictions.

Predictive Tracking Using Dynamic Models: Predictive tracking incorporates dynamic models that use the historical data of cell movement to predict future positions and behaviors. This method is useful in complex dynamic environments where cells exhibit non-linear behavior. Techniques involve:

1- Kalman Filters: These are used to predict the state of a linear dynamic system over time and are suitable for tracking cells with predictable, linear motion patterns.

2- Particle Filters: Offer a robust alternative for handling non-linear dynamics where multiple hypotheses of cell positions are maintained and updated over time.

Each methodology has its strengths and limitations and may be chosen based on the specific requirements of the research, such as the type of cells being tracked, the imaging

setup, and the expected cell behavior. Combining multiple tracking methodologies can often provide a more robust and comprehensive analysis, particularly in complex biological systems.

3.8 Cell Segmentation and Tracking Applications in Research and Medicine

Cell segmentation and tracking are pivotal techniques in modern biological research and medical diagnostics. These methodologies not only enable the detailed analysis of cellular behaviors and properties but also facilitate the monitoring of dynamic changes over time, providing insights that are critical in a range of applications.

3.8.1 Cancer Research

In oncology, cell segmentation and tracking techniques are indispensable for understanding the complex mechanisms of tumor progression and metastasis. By distinguishing and monitoring individual cells within a tumor, researchers can observe how cancer cells proliferate, invade new tissues, and respond to the microenvironment. This is particularly crucial for studying tumor heterogeneity, where different cells within the same tumor can behave differently. Moreover, tracking the migration of cancer cells helps in studying metastasis, whereby cancerous cells leave a primary tumor and initiate secondary growths elsewhere in the body. These insights are vital for developing targeted therapies that aim to inhibit tumor growth and prevent the spread of cancer cells.

3.8.2 Developmental Biology

Developmental biology benefits greatly from the nuanced application of cell segmentation and tracking, as these techniques allow for the detailed study of cell lineage and development patterns over time. By tracking how cells divide, differentiate, and contribute to the formation of tissues and organs, researchers can uncover the rules that govern normal development and identify deviations that may lead to congenital anomalies.

This includes tracking morphogenetic processes, such as gastrulation, organogenesis, and the formation of the nervous system. These detailed cellular activities provide crucial insights into the fundamental aspects of life and have implications for understanding developmental disorders.

3.8.3 Drug Development

Cell segmentation and tracking are also critical in the field of drug development. These techniques allow scientists to assess how different cells react to potential therapeutic treatments. By monitoring changes in cell morphology, viability, proliferation, and signaling in response to drug exposure, researchers can evaluate the efficacy and potential toxicity of new drugs. This is especially important in the development of chemotherapeutic agents, where the goal is to maximize tumor cell death while minimizing harm to normal cells. Automated tracking systems also facilitate high-throughput screening processes, enabling the rapid analysis of thousands of compounds, thereby accelerating the pace of drug discovery and development.

3.8.4 Regenerative Medicine

In regenerative medicine, understanding cell dynamics through segmentation and tracking is essential for developing effective therapies that replace, engineer, or regenerate human cells, tissues, or organs to restore or establish normal function. This includes monitoring stem cells and their derivatives to ensure correct differentiation and integration into existing biological systems. Tracking the behavior and fate of these cells in real-time can significantly enhance the design of scaffolds and other support structures intended to facilitate tissue regeneration and repair.

3.8.5 Immunology and Infectious Diseases

Cell tracking is fundamental in immunology, particularly in the study of how immune cells respond to pathogens. Observing the movement and interaction of cells such as T-

lymphocytes and macrophages during an immune response provides valuable insights into the mechanisms of immunity and the cellular basis of immune disorders. Additionally, tracking pathogen-infected cells can help in understanding disease progression and the effectiveness of immunotherapeutic interventions.

By integrating advanced imaging techniques and computational analyses, cell segmentation and tracking continue to open new avenues for research and therapeutic applications. These methodologies not only deepen our understanding of cellular functions and interactions but also pave the way for innovative treatments in personalized medicine and beyond.

3.9 Challenges

3.9.1 Technical Challenges in Implementing and Training AI Models

The integration of artificial intelligence (AI) and machine learning (ML) into microscopy image analysis presents numerous technical challenges. One of the primary concerns is the complexity of model development and deployment. Designing AI models that effectively interpret microscopic images requires extensive preprocessing of data, feature extraction, and the selection of appropriate algorithms. This process can be impeded by computational limitations, especially when dealing with high-resolution images that demand substantial processing power and memory.

Another significant technical challenge is the training of these models. AI systems rely heavily on large, well-annotated datasets for training. In microscopy, acquiring such datasets can be difficult due to the time-intensive nature of generating labeled data, where each image must be annotated by experts. Additionally, the quality of the training data critically impacts the model's performance, necessitating rigorous quality control processes to ensure the data's accuracy and representativeness.

3.9.2 Data Privacy and Ethical Issues Associated with Automated Decision-Making

The use of AI in microscopy also raises important data privacy and ethical issues. Microscopic image datasets often contain sensitive information, particularly when derived from human biological samples. Ensuring the confidentiality and integrity of this data is paramount, as unauthorized access or data breaches can lead to privacy violations.

Moreover, the automated nature of decision-making in AI-driven microscopy analysis poses ethical challenges. Reliance on automated systems can lead to scenarios where critical decisions are made without human oversight. This raises concerns about the transparency of AI decisions, as it can be difficult for users to understand how the AI arrived at a particular conclusion. Ensuring that these systems provide explainable outputs is essential to maintaining trust and accountability in AI applications.

3.9.3 Addressing Bias and Ensuring Reliability in AI-driven Systems

Bias in AI models is a critical issue that can undermine the reliability and fairness of automated microscopy image analysis. AI systems can inadvertently perpetuate or amplify biases present in the training data. For instance, if the data predominantly contains images from certain groups (e.g., based on race, gender, or age), the model may perform less effectively for underrepresented groups. Addressing these biases involves careful curation of diverse datasets and the implementation of algorithms designed to mitigate bias.

Ensuring the reliability of AI-driven systems in microscopy is another challenge. AI models must not only be accurate but also robust and generalizable across different settings and types of data. This requires extensive validation and testing under varied conditions to confirm that the models are stable and perform consistently. Furthermore, ongoing monitoring is necessary to detect and correct drifts in model performance over

time, which might be caused by changes in the types of images processed or shifts in the underlying biological or medical conditions being studied.

In conclusion, the integration of AI and ML into microscopy image analysis undoubtedly enhances the capabilities and efficiency of research in this field. However, addressing the technical, ethical, and bias-related challenges is crucial for advancing these technologies responsibly and effectively. By acknowledging and tackling these issues, researchers and practitioners can better harness the power of AI to contribute to significant breakthroughs in science and medicine.

3.10 Conclusion

The integration of artificial intelligence (AI) and machine learning (ML) into microscopy image analysis has dramatically advanced biological research. These technologies significantly enhance the accuracy, efficiency, and scalability of image analysis, enabling researchers to derive meaningful insights from extensive datasets. AI and ML facilitate detailed phenotypic profiling, complex morphological categorization, and dynamic bioprocess analysis, accelerating scientific discoveries and innovations.

Machine learning algorithms, such as convolutional neural networks (CNNs) and support vector machines (SVMs), have revolutionized traditional image analysis methods. Deep learning, a subset of ML characterized by deep neural networks, excels in handling high-dimensional and complex datasets. This has expanded the potential for biological discovery, particularly when combined with rigorous data preparation and preprocessing techniques like data augmentation, normalization, and meticulous labeling, ensuring the development of robust and reliable ML models.

The application of AI and ML extends to essential methodologies in microscopy, such as cell segmentation and tracking. These techniques are critical for understanding

cellular dynamics, including cell morphology, migration, division, and interaction. Advances in these areas have profound implications for fields like cancer research, developmental biology, drug development, and regenerative medicine. The ability to accurately segment and track cells enhances our understanding of cellular behavior in both normal and pathological states, providing insights that are crucial for developing targeted therapies and improving disease diagnostics.

Despite these advancements, the implementation of AI and ML in microscopy image analysis presents technical, ethical, and bias-related challenges. Addressing these issues is essential for the responsible and effective use of AI technologies. Efforts to mitigate biases, ensure data privacy, and maintain the reliability of AI-driven systems are vital for advancing the field. Overall, the integration of AI and ML into microscopy image analysis not only enhances existing methodologies but also opens new avenues for scientific discovery, ultimately contributing to significant breakthroughs in science and medicine.

4 DeepSea: An Efficient Deep Learning Model for Single-Cell Segmentation and Tracking in Time-Lapse Microscopy

4.1 Introduction

Cells frequently adapt their behavior in response to environmental cues to make important fate decisions, such as whether to divide or not. In addition, individual cells within a clonal population and under identical conditions display heterogeneity in response to environmental cues [124]. In recent years, it has become clear that single-cell level analysis over time is essential for revealing the dynamics and heterogeneity of individual cells [125, 126].

Single-cell quantitative live microscopy can directly capture both dynamics and heterogeneity of cellular decisions by continuous long-term measurements of cellular features [127, 128]. Widely available microscopy techniques such as label-free phase-contrast live microscopy allow for monitoring the dynamics of morphological features such as the size and shape of the cells [129]. The key to the successful application of single-cell live microscopy is the scalable and automated analysis of a large dataset of images. Typical live-cell imaging of biological features of cells is a multi-day experiment that produces several gigabytes of images collected from thousands of cells [127]. A major challenge for quantitative analysis of these images is the difficulty of accurately defining the borders of a cell, segmentation and tracking them over time. Low signal-to-noise ratio, existing non-cell small particles in the background, the proximity of cells, and unpredictable movements

are among the challenges for software-based automated image analysis of live single-cell microscopy data. In addition, cells are non-rigid bodies, and thus, tracking them is more challenging because they can change their shapes with time. Most critically, they divide into two new daughter cells during mitosis, which is unique and not comparable with other phenomena we encounter in conventional object-tracking applications. Solving single-cell microscopy challenges requires integrating different disciplines, such as cell biology, image processing, and machine learning.

In recent years, deep learning (DL) has outperformed conventional rule-based image processing techniques in tasks such as object segmentation and object tracking [130-132]. Traditional image segmentation approaches often require experiment-specific parameter tuning, while DL schemes are adaptive and trainable. More recently, DL-based image processing methods have attracted attention among cell biologists and microscopists, for example, to localize single molecules in super-resolution microscopy [133], enhance the resolution of fluorescence microscopy images [134], develop an automated neurite segmentation system using a large 3D anisotropic electron microscopy image dataset [135], design a model to restore a wide range of fluorescence microscopy data [136], and train a fast model that refocuses 2D fluorescence images onto 3D surfaces within the sample [137]. In particular, DL-based segmentation methods have greatly facilitated the task of cell body segmentation in microscopy images [138-141]. However, the successful application of DL-based models for time-lapse microscopy depends on applying the segmentation and tracking models in one platform to automate the analysis of a large sequence of images of live cells.

Here, we developed a versatile and trainable deep learning model for cell body segmentation and cell tracking in time-lapse phase-contrast microscopy images of mammalian cells at the single-cell level. Using this model, we developed a user-friendly

software tool, termed DeepSea, for automated and quantitative analysis of phase-contrast microscopy images. We showed that DeepSea captures dynamics and heterogeneity of cellular features such as cell cycle division and cell size in different cell types. Our analysis of cell size distribution in mouse embryonic stem cells revealed that despite their short G1 phase of the cell cycle, embryonic stem cells exhibit cell size control in the G1 phase of the cell cycle.

4.2 Methods

4.2.1 Overview of designing and training a cell segmentation and tracking model

As illustrated in Figure 4.1, first, we created an annotated dataset of phase-contrast live image sequences of three cell types: 1) mouse embryonic stem cells, 2) bronchial epithelial cells, and 3) mouse C2C12 muscle progenitor cells. To facilitate manual annotation of the cells, we developed a Matlab-based software to generate a labeled training dataset, including pairs of original cell images and corresponding cell ground-truth mask images (our annotation software is available here https://deepseas.org/software/). To further generalize our model, we used image augmentation techniques to increase the size of our dataset with more variations efficiently and less expensively. In addition to six conventional image augmentation techniques with random settings such as cropping, changing the contrast and brightness, blurring, applying the vertical/horizontal flip, and adding Gaussian noise [142, 143], we proposed and applied a random cell movement method as a novel image augmentation strategy to generate new cell images (with their annotated masks) that look more different than the original existing samples (Figure 4.2). Next, we used the annotated and augmented dataset of cell images to train our supervised DL-based segmentation model called DeepSea to detect and segment the cell bodies. To design our DeepSea segmentation model, we were inspired by the UNET model, which has been successful in different segmentation tasks [144]. We made several innovative

changes to make this model more suitable for single-cell live microscopy. First, we scaled down 2D UNET to considerably reduce the number of parameters and thus have a faster model processing large high-resolution images with less computational and memory costs. To do this, we modified our model with convolutional residual connections to increase the depth of the network with fewer extra parameters [145-147]. Second, we added an auxiliary edge detection layer trained on the edge area between touched cells to enhance the learning algorithm to focus on touching cell edges and thus improve the segmentation accuracy in hard samples with high-density touched cell images (Figure 4.3A). In the training process, we also used a progressive learning technique (used in progressive GANs [148]) to help the model generalize well for different image resolutions and generate large high-resolution masks that better separate the touching cell edges (Figure 4.4). The progressive learning technique makes the model first learn coarse-level features and then finer information.

To visualize the dynamics of cellular behavior over time, we added cell tracking capability to our DeepSea model. We trained a DL tracking model to localize and link single cells from one frame to the next and detect cell divisions (mitosis). As shown in Figure 4.3B, we used a baseline architecture similar to the DeepSea segmentation model. This model extracts the convolutional information from two consecutive image inputs (segmented cell images of times t-τ and t) to localize and detect the same target single cell or its daughter cells among the segmented cells in the current frame (time t) by generating a binary mask (Figure 4.5 and Figure 4.6). With this model, we could monitor multiple cellular phenotypes and several cell division cycles across the microscopy image sequences to generate lineage tree structures of cells. To make our model widely accessible, we developed a DL-based software with a graphical user interface (Figure 4.8) that allows researchers with no background in machine learning to automate the

measurement of cellular features of live microscopy data. We added manual editing options to DeepSea software to allow researchers to correct our model outputs when needed to bring all the DeepSea detections to the highest possible accuracy and, thus, fully track the life cycle of the cells. An interesting feature of our software is that it also allows researchers to train a new model with an annotated dataset of any cell type. We provide step-by-step instructions on how to use our software and train a model with a new dataset. Our software, instructions, and cell images dataset are publicly available at https://deepseas.org/.

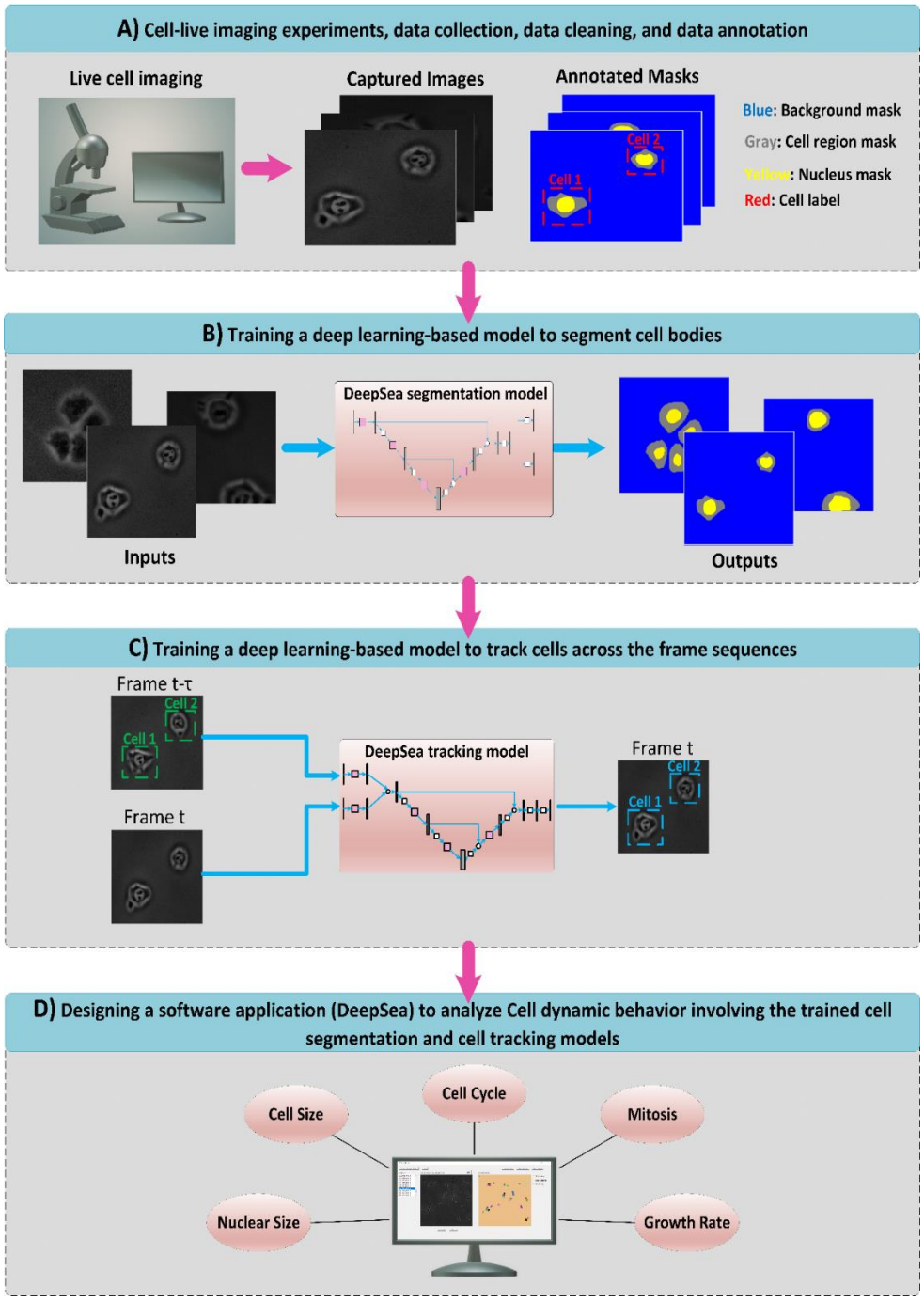

Figure 4.1: An overview of our approach. A) Generating an annotated dataset of images for training a deep learning model, B) Designing a deep learning model for segmentation, C) Designing a deep learning model for tracking single cells across sequences of phase-contrast images, and D) Developing user-friendly software to analyze cell biological features in live microscopy data.

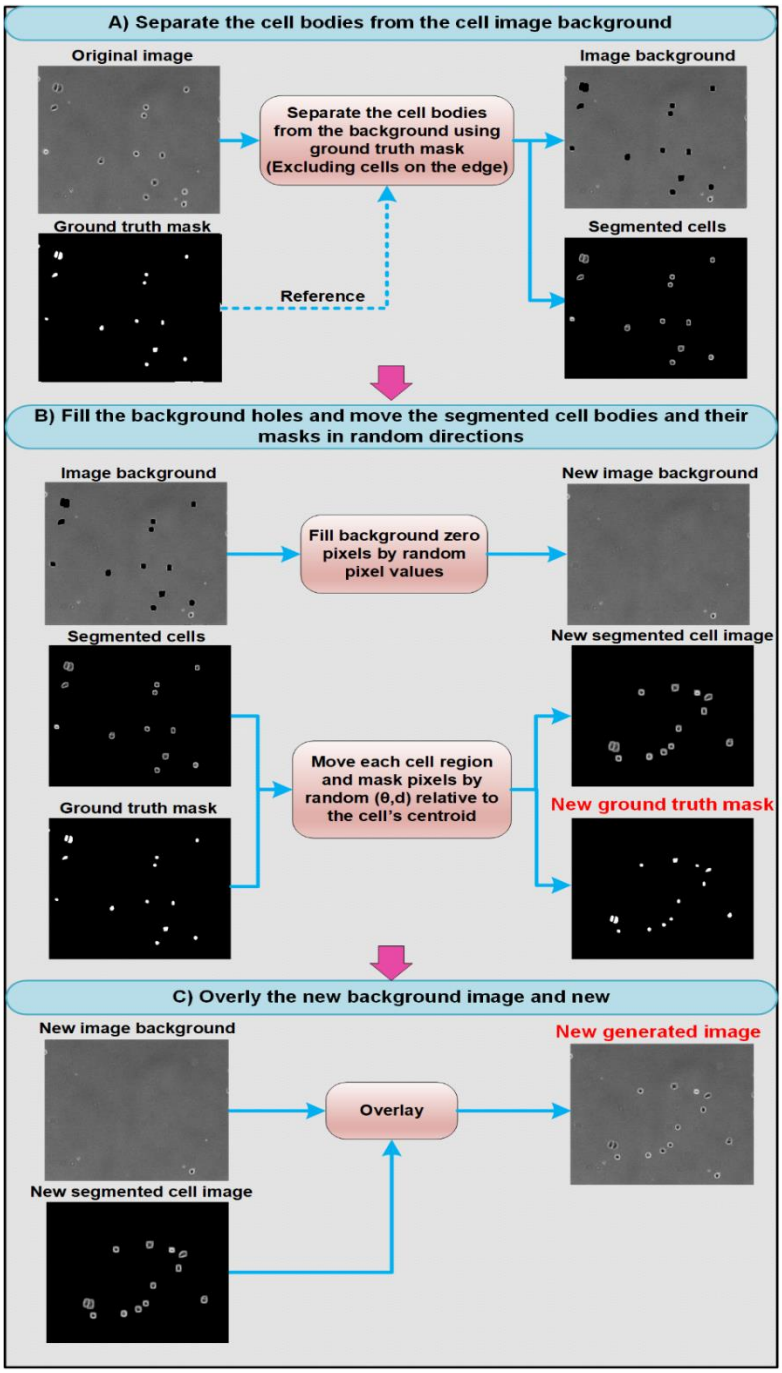

Figure 4.2: The proposed random cell movement block diagram. It can perform cell image augmentation more deeply than conventional image augmentation methods. It generates new cell images with their annotated masks from the original existing samples that look very different. θ is the direction angle between 0 and 360, and d is the displacement in pixels.

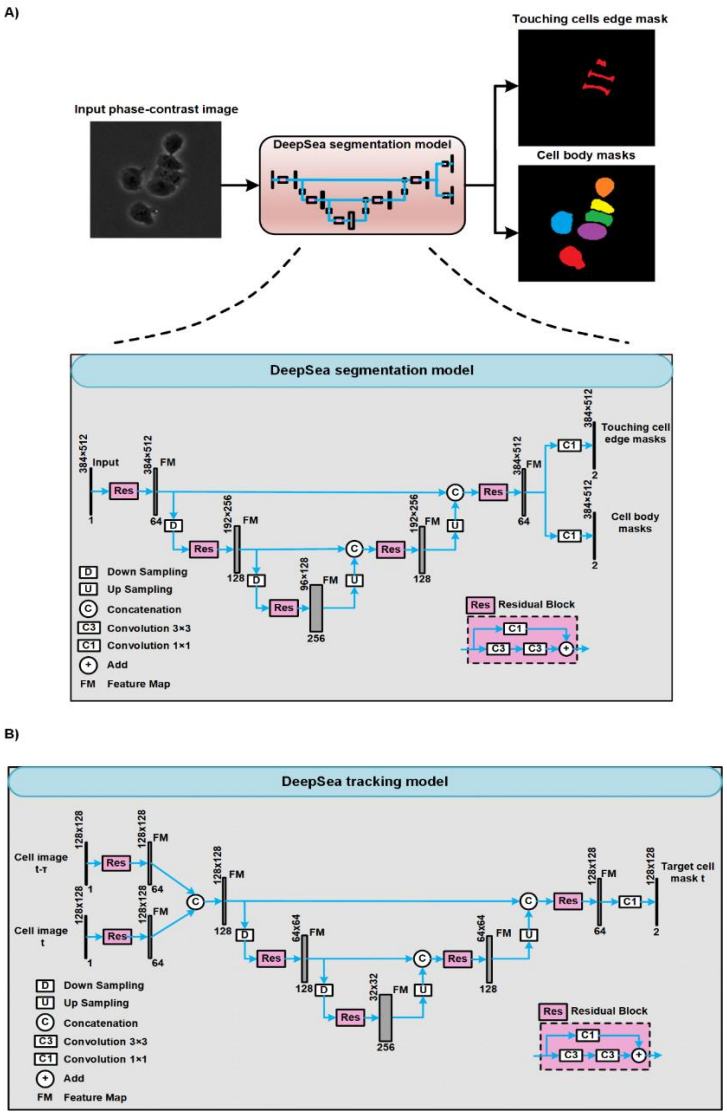

Figure 4.3: DeepSea segmentation and Tracking models. (A) The DeepSea segmentation model receives the label-free microscopy cell image and returns two outputs of the touching cell edge mask and the segmented cell body mask. This model architecture applies 1) a scaled-down version of 2D-UNET, 2) residual blocks to increase the depth of the model with fewer parameters, and 3) the auxiliary touching cell edge representations to improve the performance of the model, especially in high-density cell cultures. (B) DeepSea tracking model architecture with two input images of subsequent time points and the output of a binary mask localizing the target single cell on the current frame. Since it uses a segmentation-based process to localize and link the target cells across the frame sequences, we proposed an architecture similar to the DeepSea segmentation model as a fast and accurate enough architecture.

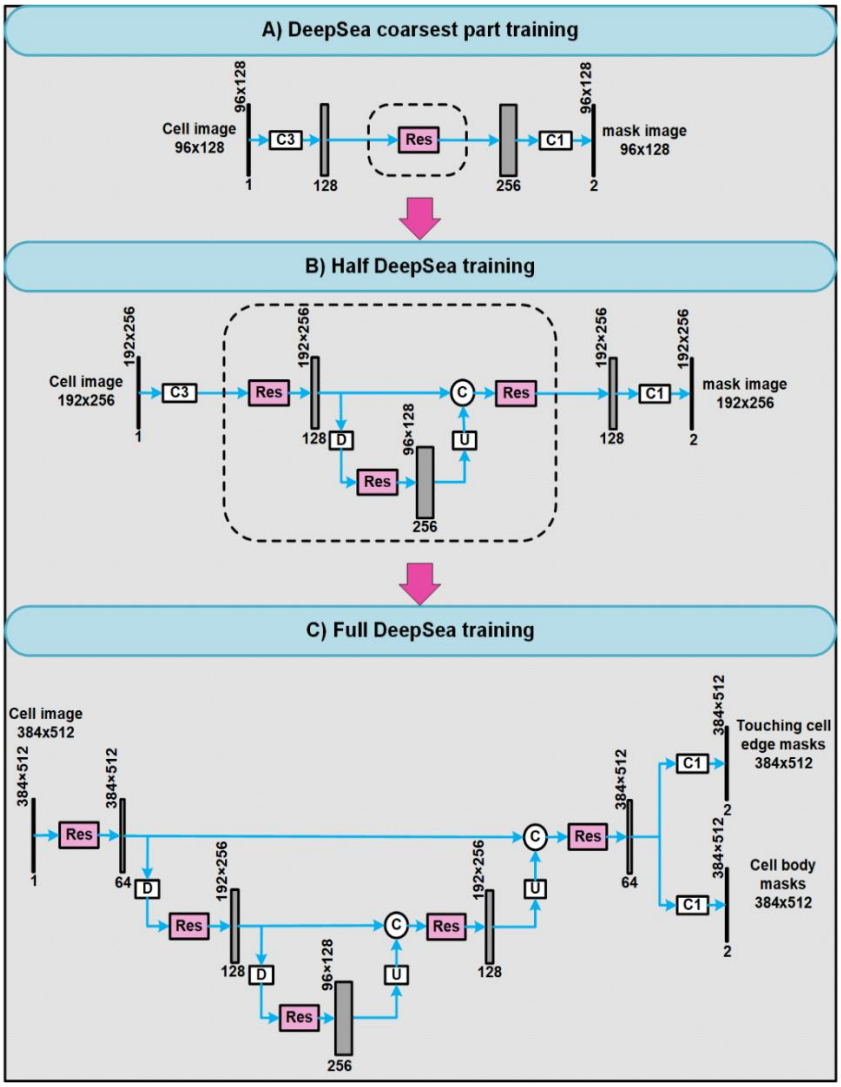

Figure 4.4: DeepSea progressive training stages. (A) First, the training algorithm starts training the coarsest part on low-resolution ground truth images of 96X128. (B) After some training epochs, it transfers the Res block weights to the half DeepSea model and keeps training it with the ground truth images of 192x256. (C) Finally, it finishes the last n training epochs with the full DeepSea model training.

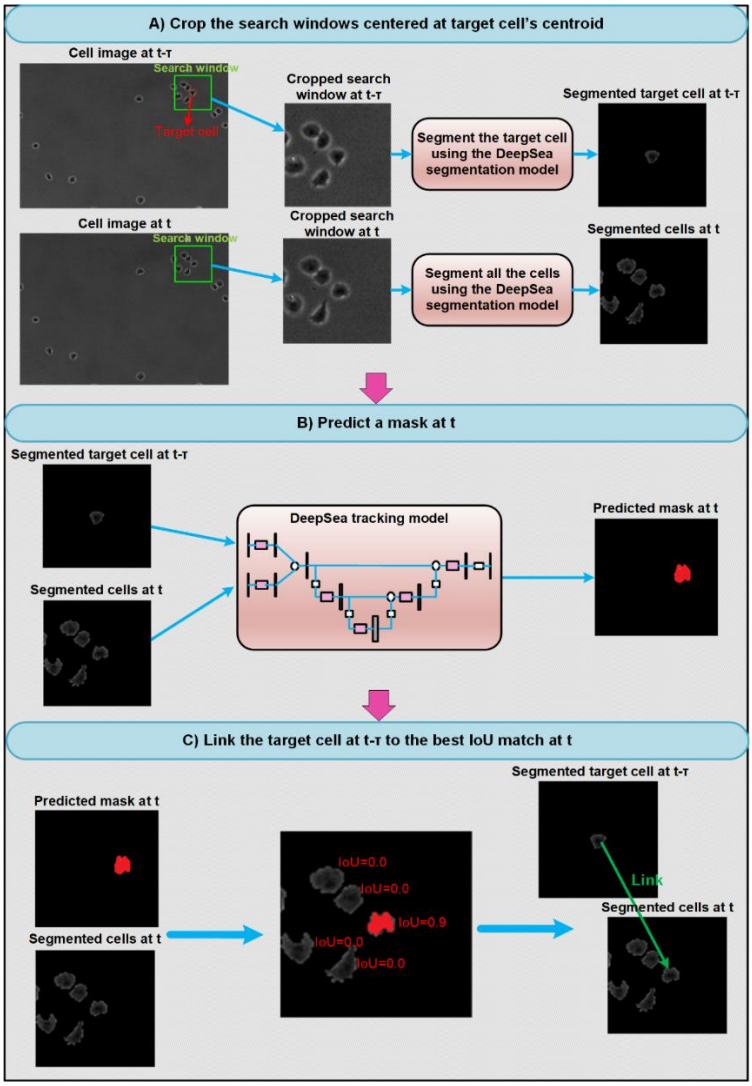

Figure 4.5: Single-cell tracking example from one frame to the next frame. (A) We limited our search space in x and y coordinates to a small square with the size of 5 times the target cell size centered at the previous frame target cell's centroids. Then, we fed each search crop into the DeepSea segmentation model to have only the segmented bodies of the target single cell in the previous frame and the segmented cells in the current frame. (B) The tracking model predicts the target single-cell location among the segmented cells on the current frame by generating a binary mask. (C) We validated the predictions using the IoU (Intersection over Union) score. We used the IoU score as a validation score to match the tracking model binary mask to each segmented cell body on the current frame and then find the true link (target cell at t-τ to selected cell at t) corresponding to the highest IoU value. A valid IoU value should be higher than a pre-defined threshold value, e.g., IoU_thr=0.5. If the model finds two or more valid IoU values, it takes it as a mitosis occurrence and thus creates the mother-daughter links between the target cell of the previous frame and the two selected cells with the highest IoU values.

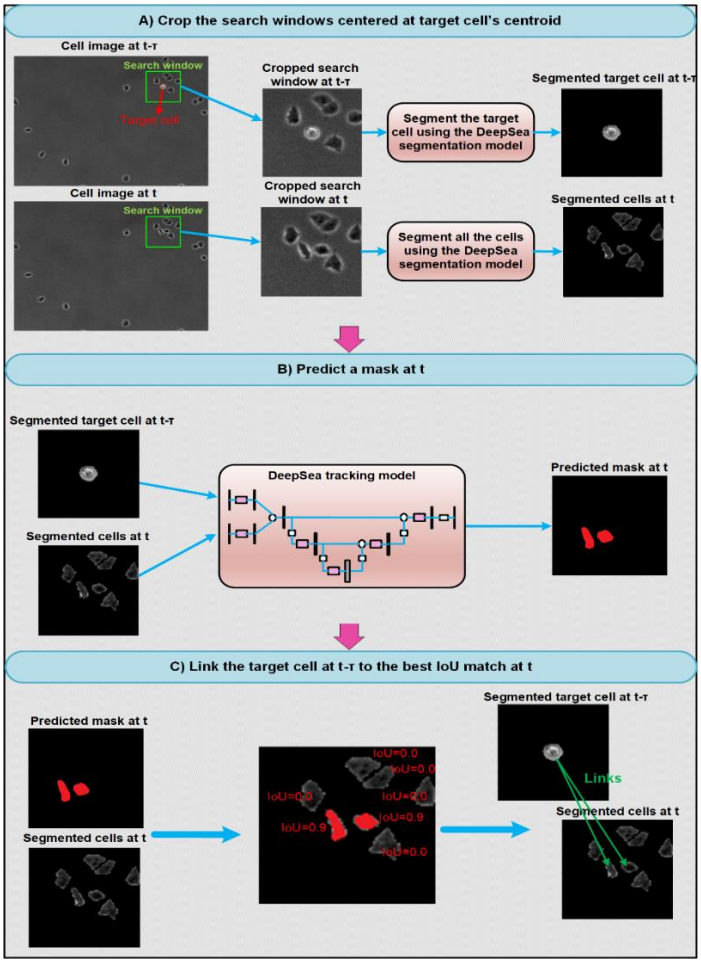

Figure 4.6: Daughter cell detection example from one frame to the next frame. (A) We limited our search space in x and y coordinates to a small square with the size of 5 times the target cell size centered at the previous frame target cell's centroids. Then, we fed each search crop into the DeepSea segmentation model to only have the segmented cell bodies of the target single cell on the previous frame and segmented cells on the current frame. (B) The tracking model predicts the daughter cell locations among the segmented cells on the current frame by generating a binary mask. (C) We validated the prediction using the IoU (Intersection over Union) score. We used the IoU as a validation score to match the tracking model binary mask to each segmented cell body on the current frame and then find the true mother-daughter links corresponding to the highest IoU values. A valid IoU value should be higher than a pre-defined threshold value, e.g., IoU_thr=0.5. If the model finds more than two valid IoU values, it creates the mother-daughter links between the mother cell of the previous frame and the two selected cells with the highest IoU values. Also, if the model finds only one valid IoU value in the current frame, it takes it as a single-cell tracking (non-mitosis) event and thus creates a single link between the target cell of the previous frame and the single valid prediction on the current frame.

4.2.2 Cell Culture and Microscopy

Mouse ESCs (V6.5) were maintained on 0.1% gelatin-coated cell culture dishes in 2i media (Millipore Sigma, SF016-100) supplemented with 100U/ml Penicillin-Streptomycin (Thermo Fisher, 15140122). Cells were passaged every 3-4 days using Accutase (Innovate Cell Technologies, AT104) and seeded at a density of 5,000-10,000 cells/cm2. For live imaging, between 5000 to 10,000 cells were seeded on 35mm dishes with a laminin-coated (Biolamina) 14mm glass microwell (MatTek, P35G-1.5-14-C). Cells were imaged in a chamber at 37C perfused with 5% CO2, a Zeiss AxioVert 200M microscope with an automated stage, and an EC Plan-Neofluar 5x/0.16NA Ph1objective or an A-plan 10x/0.25NA Ph1 objective. The same culture condition was used for confocal imaging, except that 24 hours after seeding, the media was replaced with 2ml DMEM-F12 (Thermo Fisher, 11039047) containing 2ul CellTracker Green CMFDA dye (Thermo Fisher, C2925) and placed back in the incubator for 35 minutes. Next, 2 ul of CellMask Orange plasma membrane stain (Thermo Fisher, C10045) was added, and the dish was incubated for another 10 minutes. Dishes were washed three times with DMEM-F12, after which 2ml of fresh 2i media was added. Cells were imaged directly after the live-cell staining protocol using the Zeiss 880 Microscope using a 20x/0.4 N.A. objective and a 1μm interval through the z-axis.

Immortalized human bronchial epithelial (HBEC3kt) cell line homozygous for wildtype U2AF1 at the endogenous locus was obtained as a gift from the laboratory of Harold Varmus (Cancer Biology Section, Cancer Genetics Branch, National Human Genome Research Institute, Bethesda, United States of America and Department of Medicine, Meyer Cancer Center, Weill Cornell Medicine, New York, United States of America) and cultured according to Fei et al. [149]. This host cell line was used for lentiviral

transduction and blasticidin selection to generate a line with stable expression of $KRAS^{G12V}$ using a lentiviral plasmid obtained as a gift from the laboratory of John D Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center) described in [150]. Cells from passage 11 were grown to 80% confluency in Keratinocyte SFM (1X) (Thermo Fisher Scientific, USA) before being re-seeded as biological duplicates at three densities: 0.3M, 0.2M, and 0.5M cells per well in 6-well plates and allowed to adhere before live-cell imaging over a 48h time period.

4.2.3 Dataset

We collected phase-contrast time-lapse microscopy image sequences of three different cell types, including two in-house datasets of Mouse Embryonic Stem Cells (MESC, 31 sets, 1074 images) and Bronchial epithelial cells (7 sets, 2010 images) and one dataset of Mouse C2C12 Muscle Progenitor Cells (7 sets, 540 images) obtained from an external resource with the cell culture described in [151]. Our collected datasets are publicly available at https://deepseas.org/datasets/. Some dataset statistics are shown in Table 4.1. We designed an annotation software in MATLAB (https://deepseas.org/software/) to manually create the ground-truth mask images corresponding to our cell images. We applied an image augmentation scheme to generate a larger dataset with more variations efficiently and less expensively, aiming to train a more generalized model. In our image augmentation scheme, in addition to conventional image transformations [142, 143], we proposed moving the stem cell bodies by the random vectors of (θ,d) relative to their center points, where θ is the direction angle between 0 and 360 and d is the displacement in pixels (Figure 4.2). The proposed cell image augmentation method improved the model performance with unseen test images (different microscopy live imaging sets not used in the training set), confirming that it could less overfit training

samples and thus help the model generalization. For each training image, we applied a pipeline of augmentation functions, which were randomly selected and set.

Table 4.1: DeepSea dataset characteristics.

	Image Size	Frame Rate	Num of Sets	Num of Images	Num of Single Cells	Num of Cell Cycles
Stem Cells	1344x1024	15-30 min	30	2010	14995	115
Bronchial Cells	1244x904	5 min	8	1174	48027	292
Muscle Cells	1392x1040	20 min	9	502	22080	274

4.2.4 Segmentation model

As mentioned before, our dataset samples are label-free microscopy images that are usually noisy, low contrast, hard, and high cell density samples. It is difficult for any existing instance tools (that have not seen these types of images in their training process) to segment the cell bodies of our test images. The original pre-trained version of StraDist and StarDist models achieved an average precision of around 43% and 5%, respectively, on our test sets. Figure 4.7 shows the CellPose and StarDist outputs compared with the ground truth mask images.

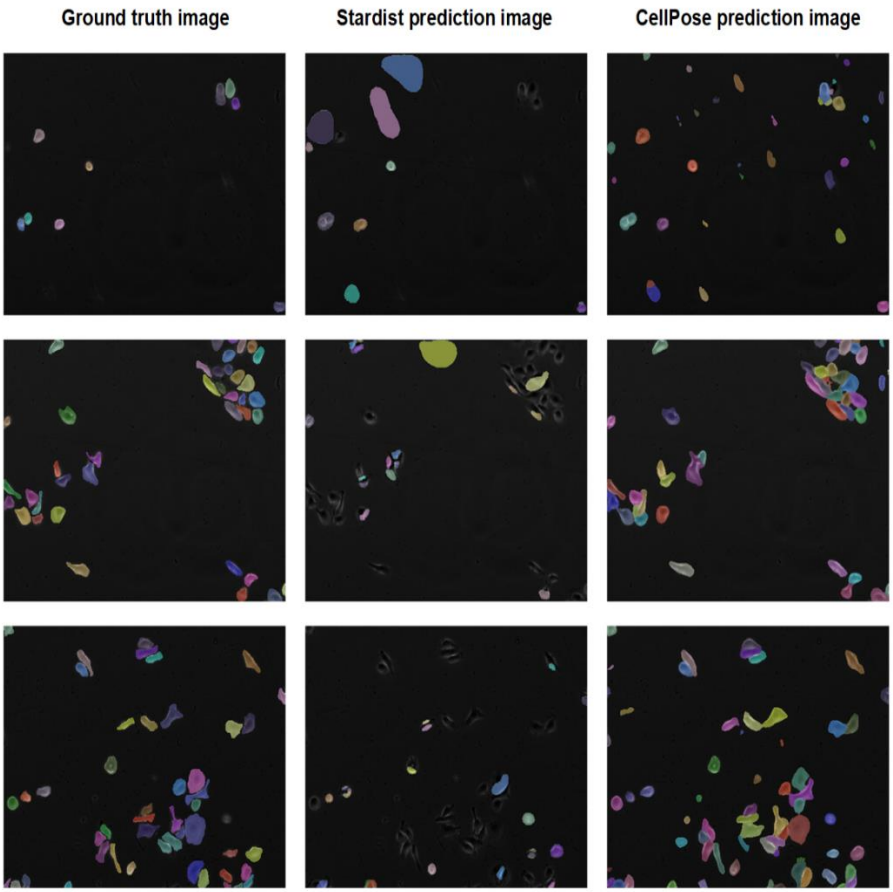

Figure 4.7: Comparative performance analyses. The original pre-trained CellPose and StarDist model outputs compared with the ground truth cell body masks.

In the instance segmentation task, we proposed and built a 2D deep learning-based model called DeepSea (Figure 4.3A). To design our DeepSea segmentation model, we were inspired by the UNET model [144, 152]. Since we needed a fast segmentation and tracking model to be used in our DeepSea software, we decided to reduce the number of parameters and make a scaled-down version of 2D UNET. By reducing the model size, we could feed larger high-resolution images into the model and get more accurate results [153, 154] with less computational and memory costs. However, to compensate for the model compression and also avoid the model from underfitting the training data, we

modified the scaled-down 2D-UNET model with the convolutional residual connections. It has been proved that the residual connections can increase the depth of the network with fewer extra parameters. They can also accelerate the speed of training the deep network, reduce the effect of the vanishing Gradient Problem, and potentially obtain higher accuracy in network performance [145, 147]. Our DeepSea segmentation model involves only 1.9 million parameters, which is considerably smaller than typical instance segmentation models such as UNET [152], PSPNET [155], and SEGNET [156].

During the training process, we started training the model with the low-resolution images 95x128, then increased it to 191x256, and finished it with 384x512, as described in Figure 4.4. Our learning algorithm started with the lowest resolution part and then progressively added the other high-resolution blocks until the desired image size and full DeepSea model were achieved. The progressive learning technique (as used in progressive GANs [148]) can help the model generalize well for different image resolutions and generate large-high-resolution masks that better separate the touching cell edges. Also, when adding the higher resolution part to the training process, our learning algorithm reduces the learning rate of previously trained parts, making the different parts of the model learn information from different resolutions independently.

The auxiliary edge representations (highlighting the edge area between touching cells) and the auxiliary training loss value (Equation 4.2) also encouraged the learning algorithm to spend more computational budget and time to separate the touching cells. They thus improved the model performance, especially for hard samples where we have high-density touching cells. We also artificially increased and repeated the hard cell images in our training dataset to make the model see them more during the training process (almost the same number as non-touching samples). This also helps the learning algorithm

balance the loss functions (Equation 4.2). To create each touching cell edge mask, we first created a weight map from the ground truth cell masks according to Equation 4.1:

$$w(x) = w_{0.} \exp(-\frac{(d_1(x)+d_2(x))^2}{2\sigma^2}) \quad (4.1)$$

, where x is pixels in the image, d1 is the distance to the border of the nearest cell, d2 is the distance to the border of the second nearest cell, and w_0 and σ were set to 10 and 25, respectively. Then, we make a binary image by replacing all pixel values above a determined threshold (=1.0) with 1s and setting all other pixels to 0s.

In the training process, we used the early stopping technique to stop training when the validation score stopped improving. We also took advantage of batch normalization and dropout techniques to improve the model's speed, performance, and stability [157]. Besides, the image augmentation pipeline we designed (Figure 4.2) could help the model see more variations during the training process and then process the unseen test samples more confidently. We chose the RMSprop optimization function with the learning rate scheduler of the OneCycleLR method (LR=1e-3) to optimize model weights and minimize the proposed loss function (Equation 4.2). Our loss function is a linear combination of cross-entropy (CE) loss and Dice loss (DL) functions [158], as well as auxiliary loss functions (EdgeCE and EdgeDL) for the touching cell edge representations. CE takes care of pixel-wise prediction accuracy, while DL helps the learning algorithm increase the overlap between true area and predicted area, which is essentially needed where the number of image background pixels is much higher than foreground pixels (object area pixels).

$$\text{Loss} = \text{CE} + \text{DL} + \text{EdgeCE} + \text{EdgeDL} \quad (4.2)$$

In the test phase, we used the IoU index, a value between 0 and 1 and known as the Jaccard index as well [159] (Equation 4.3), to match the segmentation model predictions to the ground truth annotated masks:

$$\text{IOU} = \frac{\text{Area of overlap between predicted pixels and ground truth pixels}}{\text{Area of union encompassed by both predicted pixels and ground truth pixels}} \tag{4.3}$$

In each test image, we labeled each detected cell body whose IoU index was higher than a pre-defined threshold value as a valid match and a True Positive (TP) prediction. Also, the ground truth cell body masks with no valid match were categorized into the False Negative (FN) set, and the predictions with no valid ground truth masks were labeled as the False Positive (FP) cases (non-cell objects). Then, using Equation 4.4, we calculated the average precision (AP) value for each image in the test set used by the other state-of-the-art methods in cell body segmentation tasks [138]:

$$\text{AP} = \frac{\text{TP}}{\text{TP+FN+FP}} \tag{4.4}$$

4.2.5 Tracking model

Our tracking model aimed to localize and link the same target single-cell bodies from one frame to the next and also detect cell divisions (mitosis). We used a baseline architecture similar to the DeepSea segmentation model (as a fast and accurate enough architecture) but with multiple images, two inputs, and one output (Figures 4.3B, 4.5, and 4.6). The first input is the target cell image at the previous frame (previous time point t-τ), the second input is the segmented cell image at the current frame (current time point t), and the output is a binary mask at the current frame. This model extracts the convolutional information from the input images to localize and find the target single cell or its daughter cells among the segmented cells on the current frame by generating a binary mask (Figures 4.5B and 4.6B). To increase the accuracy of the tracking model, we limited our

search space in x and y coordinates to a small square with the size of 5 times the target cell size centered at the previous frame target cell's centroids (Figures 4.5A and 4.6A). Since cells move slowly through space, the cell's previous location presents a good guess of where the model should expect to find it in the current frame. To validate the tracking model's output, we used the IoU (Intersection over Union, Equation 4.3) as a validation score. We matched the tracking model binary mask to each segmented cell body on the current frame and measured the IoU value (Figures 4.5C and 4.6C). In the validation process, if the IoU score of a segmented cell body on the current frame was higher than a pre-specified IoU matching threshold value (e.g., IoU_thr=0.5), we labeled it as a positive detection and valid link. Then, we categorized them into true or false positive detections by comparing them with ground truth cell labels aiming to measure the average precision (AP metric introduced in Equation 4.4) of the tracking model in tracking the target cell bodies from one frame to the next and detecting cell divisions.

The number of the DeepSea tracking model parameter is only 2.1 million, while the other deep tracking models, such as ROLO [160], DeepSort [161], and TrackRCNN [162], which are mostly used in other object-tracking applications, involve more than 20 million parameters, confirming that we have an efficient model in the tracking process as well. Also, since the number of cell division events is naturally much fewer than single-cell tracking events, we artificially repeated and increased the cell division events fifty times more than single-cell tracking events in our training set. This helped the model see a balanced number of both single-cell links and cell divisions during the training process and thus reduced the risk of overfitting the most repeated category. The train optimization function and hyperparameters are the same as the segmentation model training process.

To evaluate our tracking model in a continuous cell trajectory tracking process during an entire cell life cycle from birth to division, we used MOTA (Multiple Object Tracking

Accuracy, Equation 4.5), which is widely used in multi-object tracking challenges [130, 163, 164]. To our knowledge, this is the first time that this metric has been used to evaluate a cell tracking model performance. We also used other commonly used tracking metrics, as follows, to give more detailed evaluation information.

IDS: Identity Switch is the number of times a cell is assigned a new label in its track.

MT: Mostly Tracked is the number of target cells assigned the same label for at least 80% of the video frames.

ML: Mostly Lost is the number of target cells assigned the same label for at most 20% of the video frames.

Frag: Fragmentation is the number of times a cell is lost in a frame but then redetected in a future frame (fragmenting the track).

$$\mathrm{MOTA} = 1 - \frac{\sum_n(\mathrm{FP}_n + \mathrm{FN}_n + \mathrm{IDS})}{\sum_n(\mathrm{Numberofcells})} \tag{4.5}$$

, where n is the frame number. A perfect tracking model achieves MOTA=1.

4.2.6 Designed software tools

We designed two software tools for the Deepsea project, including 1) Manual annotation software and 2) DeepSea cell segmentation and tracking software. The step-by-step instructions with examples of how to use them are uploaded to the page at https://deepseas.org/software/. The manual annotation software is a MATLAB-based tool that we designed and used to manually segment and label the cells of the raw dataset of microscopy images we collected. This tool helped us provide the required ground truth dataset that we needed for training the cell segmentation and tracking models. It can also be used to manually annotate any other image datasets.

Also, DeepSea software (Figure 4.8) is a user-friendly and automated software designed to enable researchers to 1) load and explore their phase-contrast cell images in a high-contrast display, 2) detect and localize cell bodies using the pre-trained DeepSea segmentation model, 3) track and label cell lineages across the frame sequences using the pre-trained DeepSea tracking model, 4) manually correct the DeepSea models' outputs using user-friendly editing options, 5) train a new model with a new cell type dataset if needed, 6) save the results and cell label and feature reports on the local system. It employs our latest trained DeepSea models in the segmentation and tracking processes.

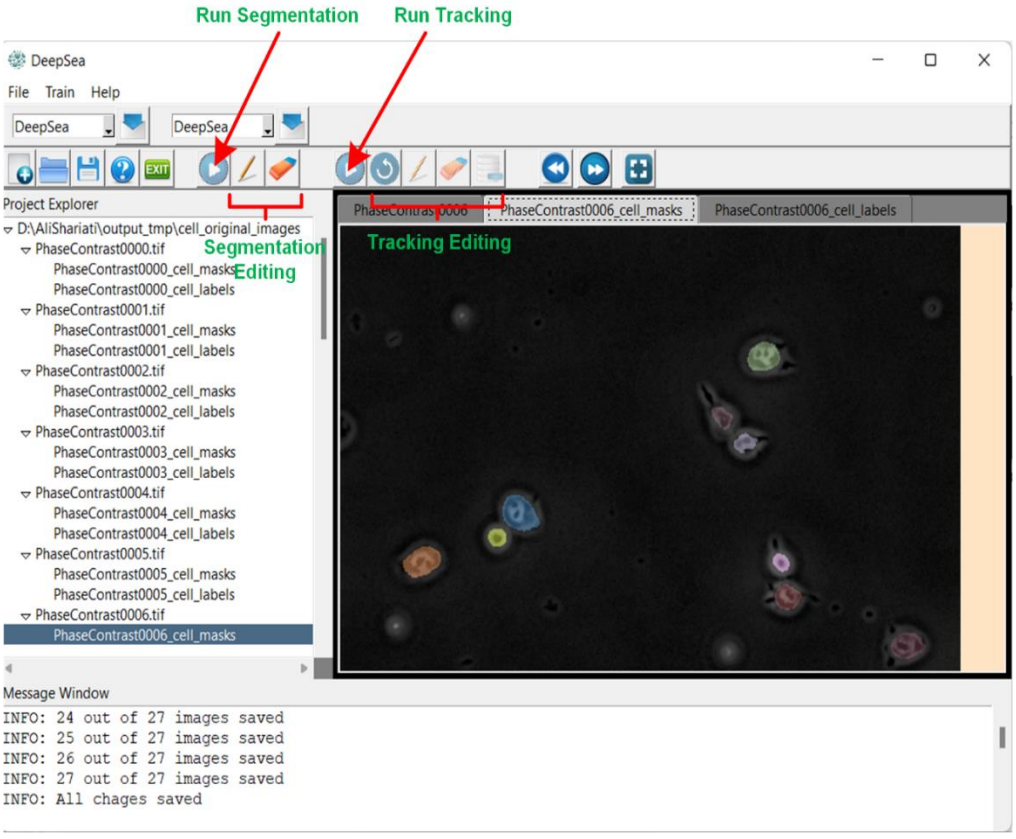

Figure 4.8: DeepSea software snapshot. It is an automated cell segmentation and tracking software that employs our latest trained DeepSea models and also provides users the editing options to manually correct the DeepSea outputs.

4.3 Results

4.3.1 Model and training configuration effects

Table 4.2 shows how our proposed techniques and modifications can improve the segmentation scores for the simple and crowded samples as measured by precision.

Table 4.2: Model and training configuration.

	Easy samples ap@0.5	Hard samples ap@0.5	Number of parameters
2D UNET Scaled Down	0.87±0.1	0.82±0.2	1035778
Modified 2D UNET Scaled Down + Res Connections	0.93±0.1	0.84±0.1	1938306
Modified 2D UNET Scaled Down + Res Connections + Progressive Learning	0.94±0.2	0.87±0.2	1938306
Our DeepSea: Modified 2D UNET Scaled Down + Res Connections + Progressive Learning + EdgeDetectionLayer/Loss	0.93±0.2	0.90±0.1	1938436

4.3.2 DeepSea performance evaluation

The trained segmentation model fits the exact boundary of the target cells and labels their pixels with different colors, helping to determine each cell's shape and area within the input microscopy image (Figure 4.3A). To evaluate the performance of our segmentation model, we compared the model's predictions to true manually segmented cell bodies at different thresholds of the standard intersection over union metric (IoU) on the test images. Next, we used the standard average precision metric, which is commonly used in pixel-wise segmentation and object detection tasks, to compare DeepSea with recently developed segmentation models. DeepSea was able to outperform existing state-of-the-art models such as CellPose [138], StarDist [139], and 2D-UNET [144, 152] in terms of latency and mean average precision (mAP) when trained on the same training sets and tested on the same test sets at all pre-defined IoU thresholds (Figure 4.9A, B). Notably, we

observed close prediction accuracy between images with a higher density of cells with touching edges (hard cases) and images with a lower density of cells (easy) with an overall higher precision compared to the CellPose model (Figure 4.9C, D). Examples of DeepSea's accuracy in high-density cell cultures are shown in Figure 4.10. In addition, we demonstrated the generalizability of the DeepSea model performance with different cell-type test images of our dataset (Figure 4.9E). Three examples of the DeepSea and CellPose segmentation model's output are compared in Figure 4.11. Next, we compared the performance of DeepSea with CellPose in measuring cellular phenotypes such as cell size. A comparison of cell size distribution obtained from DeepSea and CellPose showed that DeepSea obtains a median cell size that is closer to the median cell size obtained by manual segmentation (Figure 4.12). Together, these results indicate that DeepSea's segmentation model works robustly across different densities of cells and different cell types in our dataset with high precision.

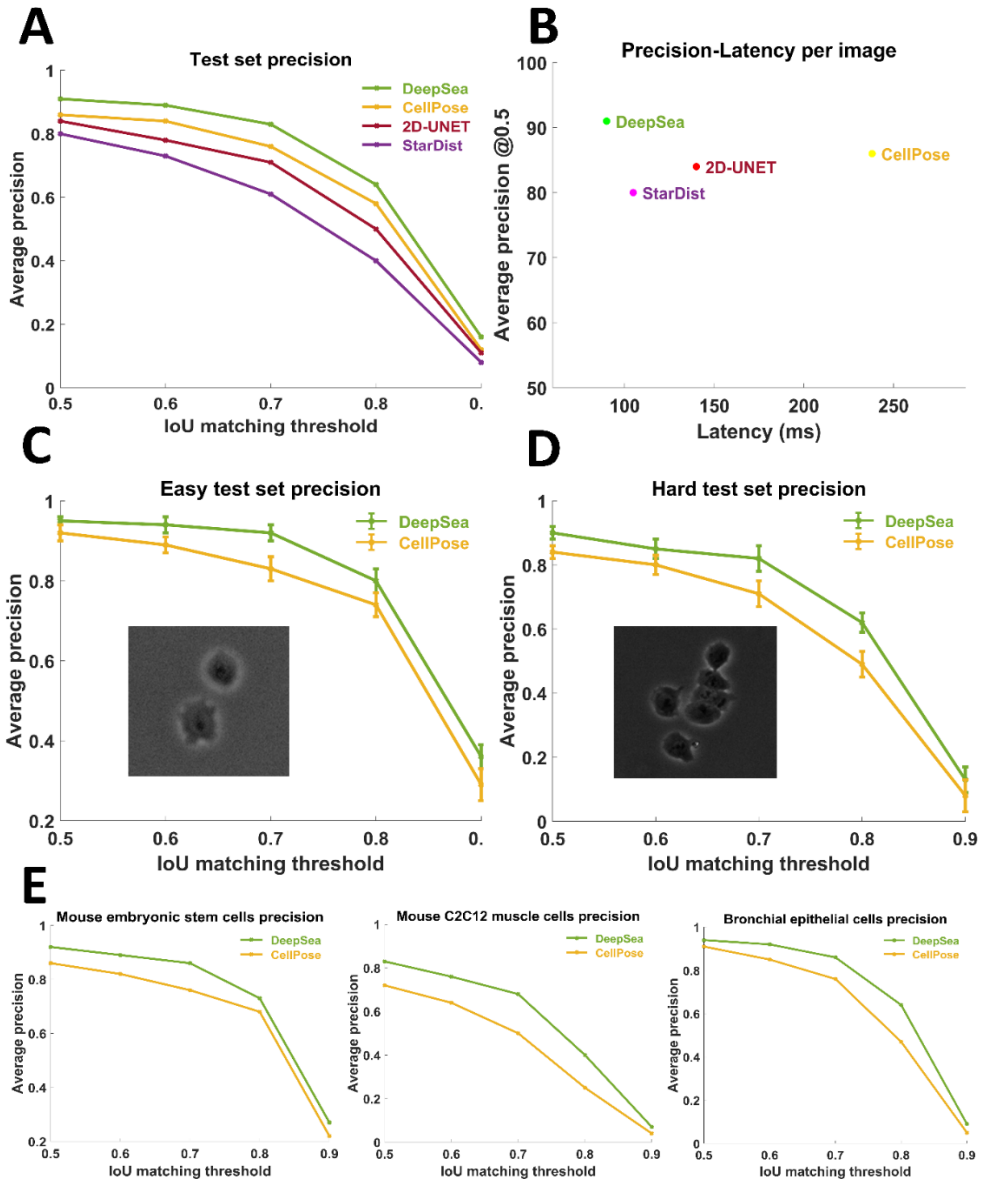

Figure 4.9: Segmentation model evaluation on the test set images. (A) Comparing the performance of DeepSea, Cell Pose, StarDist, and 2D-UNET using the standard average precision at different IoU matching thresholds. (B) Measuring models' latency (per image) to compare the DeepSea efficiency with the other models. (C-D) Comparing models' performance in segmenting easy (sparse cell density) and hard (high cell density) test images. (E) Comparing models' performance in segmenting different cell types of the DeepSea dataset.

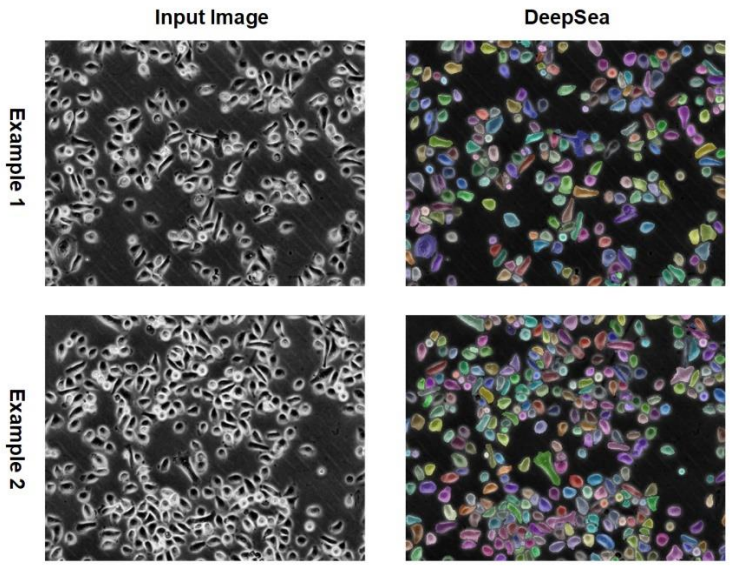

Figure 4.10: Examples of outputs. Two examples of the DeepSea segmentation model output, confirming high precision with high-density cell images.

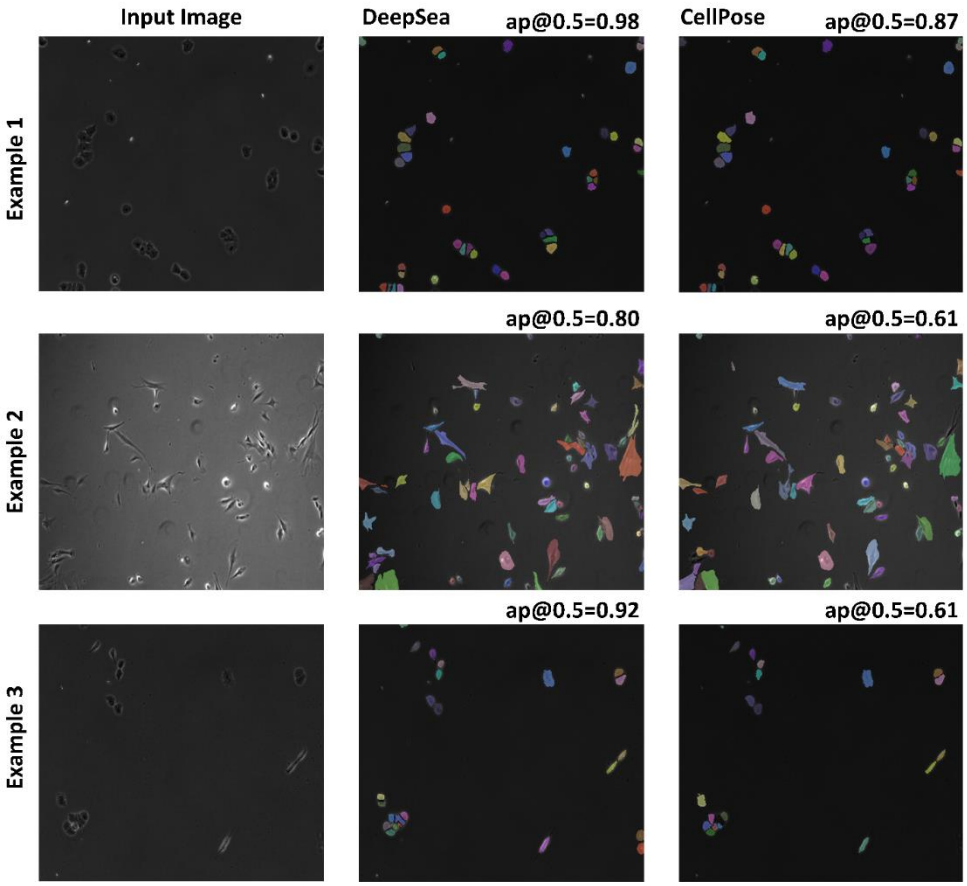

Figure 4.11: Three examples of segmentation outputs. DeepSea output (middle column) compared with the CellPose (right column) for different cell types. DeepSea has higher average precision (ap) compared to the CellPose model.

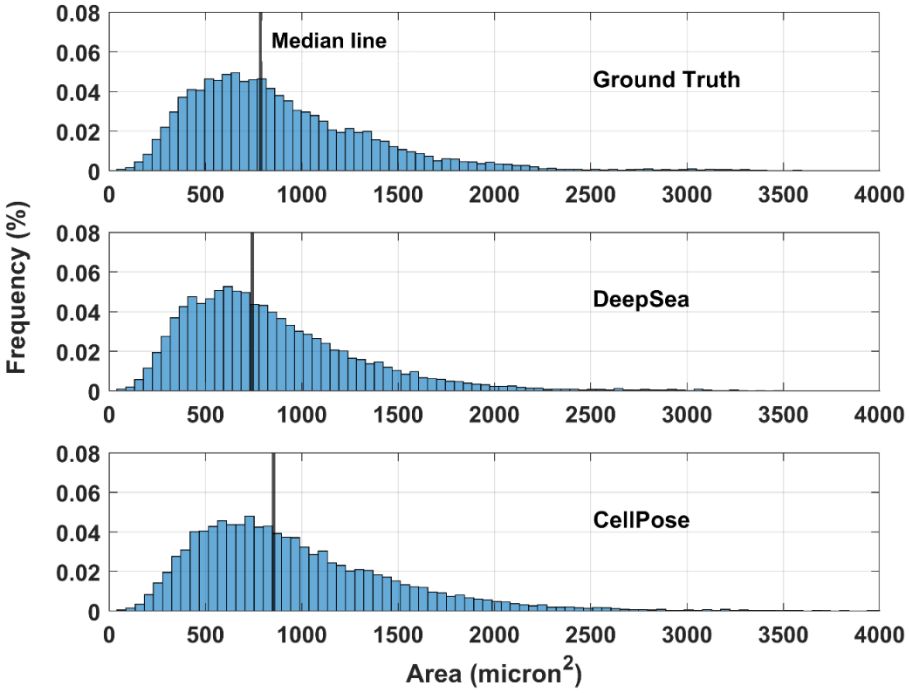

Figure 4.12: Cell size distribution of ground truth test dataset compared with the distribution obtained from DeepSea and CellPose detections.

The DeepSea tracking model receives the segmented target cell image at the previous time point and the segmented cell image at the current time point to generate a binary mask localizing the target cell (or its daughter cells) at the current time point (Figures 4.5 and 4.6). For the tracking model, we evaluated the model's performance on the test set by measuring the average precision of single-cell tracking from one frame to the next frame, as well as mitosis detections. We matched the binary masks obtained from the tracking model at time t to the true target cell bodies (at time t) at different matching thresholds of IoU. While our model achieved 0.98±0.2 precision (@0.5 IoU threshold) for tracking single cells, the precision of our model for mitosis detection was around 0.89±0.3 (@0.5 IoU threshold) (Figure 4.13A, B). Mitosis detection was particularly more challenging for stem cell images (Figure 4.14). We speculated that there might be a direct relationship

between the single-cell and mitosis detection results and the frame imaging intervals. Thus, we ran an experiment to measure the tracking model sensitivity to the frame sampling rate. We used the test frame sequences of Bronchial epithelial cells and down-sampled the frames to make the sub-sampling intervals of 5, 10, and 15 minutes. The results are shown in Figure 4.15, confirming that the model precision is sensitive to the changes in the frame's time distance. It shows that the higher sampling rate reduces the tracking model failures, especially for the cells that move and change fast over time.

Next, we systematically compared DeepSea tracking precision with some existing cell tracking tools (Table 4.3). As shown, some of these tools only support a part of the required process, either single-cell tracking [165, 166] or mitosis detection [167], and some of them are proposed to be used for both, like Trackmate [168]. Similar to the DeepSea tracking pipeline, they all first need to detect and segment the cell bodies before starting the cell tracking process and frame-by-frame cell-linking. The segmentation precision of all of them with our cell images is lower than 50%. Thus, we decided to use DeepSea segmentation outputs as the input for these tracking tools to obtain the best possible tracking results and then compare only their cell tracking part performance. We assessed the tracking model of DeepSea and other tracking tools in a full cell cycle tracking task. This test uses the trained tracking model to track and label the target single-cell motion trajectories across the live-cell microscopy frame sequences from birth to division. In this evaluation process, we used MOTA (Multi-Object Tracking Accuracy) which is a widely used metric in multi-object tracking schemes and measures the precision of localizing objects over time across the frame sequences (Equation 4.5). We also included other commonly used tracking metrics such as IDS (Identity Switch), MT (Mostly Tracked), ML (Mostly Lost), and Frag (Fragmentation) to provide more detailed evaluation information [130, 163, 164].

Our DeepSea tracking model achieved a MOTA value of 0.94±0.2 compared with the Trackmate model, with a MOTA of 0.29±0.7 (Table 4.4). In the evaluation process, we used 228 full ground truth cell cycle trajectories, each including more than three consecutive frames. The Trackmate algorithm [168] is one of the widely used cell tracking tools. The main factor for Trackmate's overall low MOTA was that it frequently did not detect mitotic events, leading to high false positive (FP) and false negative (FN) labels (Table 4.4 and Equation 4.5). We also would like to note that rule-based tools like Trackmate are not trainable and cannot be rapidly adapted to any specialized dataset.

Figure 4.16 shows one example of our model output with MOTA=1.0, tracking the cell motion trajectories over nine consecutive frames for three target cells. We also added the manual editing option to DeepSea software so that users can edit the software's output and reduce the segmentation and tracking errors. Thus, it allows researchers to fully and accurately track the life cycle of a cell.

Table 4.3: Single-cell tracking and mitosis detection precision.

	Single-cell tracking	**Mitosis detection**
Trackmate [31]	0.76	0.36
CellTracker [28]	0.69	Not supported
MDMLM [30]	Not supported	0.85
CellTracking [29]	0.82	Not supported
DeepSea Tracker	0.98	0.89

Table 4.4: Multi-cell cycle tracking results.

	MOTA	**MT**	**ML**	**Precision**	**Recall**	**Frag**	**IDS**	**FP**	**FN**	**Latency Per image (GPU NVIDIA RTX2080)**
Trackmate [31]	0.29	0.15	0.56	0.42	0.51	25	8	2614	2053	85 ms
DeepSea Tracker	0.94	0.93	0.01	0.98	0.97	62	82	153	162	580 ms

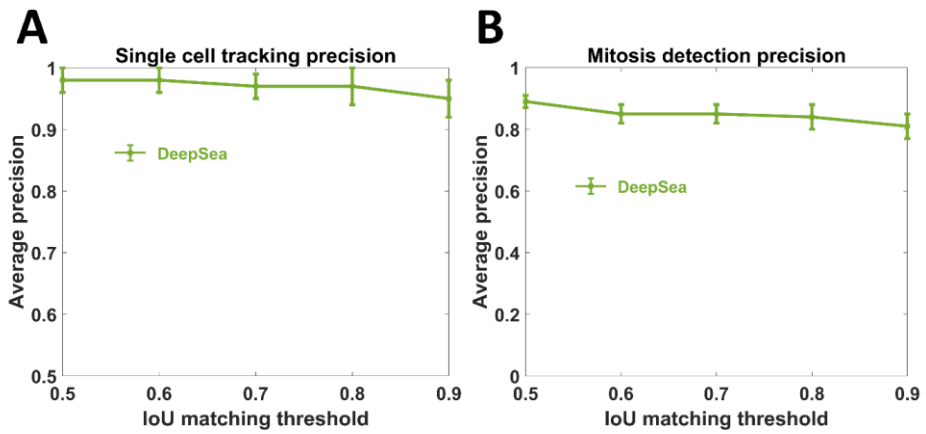

Figure 4.13: Tracking model evaluation on the test set. We evaluated the model performance using the standard average precision at different IoU matching thresholds**. A**) Single-cell tracking precision at different IoU matching thresholds. **B**) Mitosis detection precision at different IoU matching thresholds.

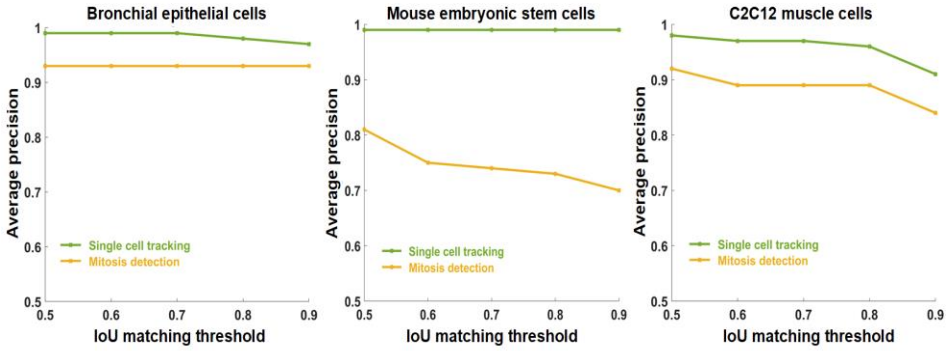

Figure 4.14: The DeepSea tracker model and mitotic detection performance with different cell types of the test set.

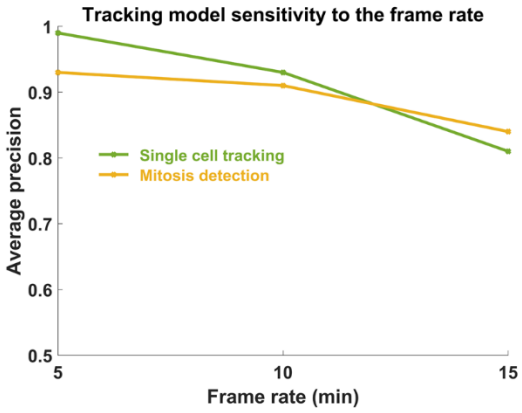

Figure 4.15: The tracking model sensitivity to the frame sampling rate. Bronchial epithelial cell images are used for this experiment.

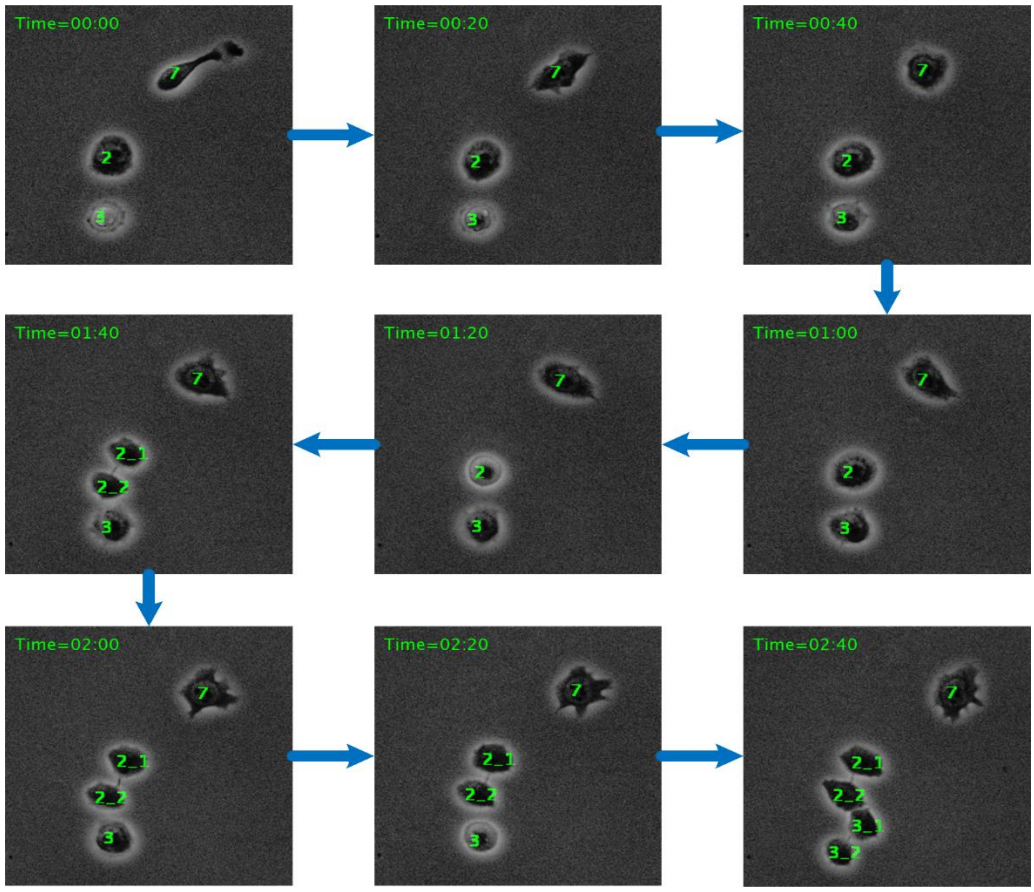

Figure 4.16: Example of the cell cycle tracking process obtained by feeding nine consecutive stem cell frames (with a sampling time of 20 minutes) to our trained tracking model. Daughter cells are linked to their mother cells by an underline (in the sixth and seventh frames).

4.4 Conclusion

Here, we introduced DeepSea, an efficient deep-learning model for automated analysis of time-lapse images of cells. The segmentation and tracking of cell bodies and subcellular organelles from microscopy images are critical steps for nearly all microscopy-based biological analysis applications. Although phase-contrast microscopy is a non-invasive and widely used method for live-cell imaging, developing automated segmentation and tracking algorithms remains challenging. Segmentation of phase-contrast images

remains difficult because of the presence of bright light artifacts such as halo at the edges of the cells and inhomogeneity in refractive index, producing noisy images. Although not unique to phase-contrast microscopy, cell tracking has its own challenges due to the unpredictable nature of cells in their movements over time, the close proximity of cells, and the division of cells.

Here, we leveraged the recent advancement in deep learning-based image processing to address some of these challenges. The lack of a comprehensive, high-quality annotated dataset of cells prevents the full utilization of deep learning-based models for microscopy image analysis systems. We generated large manually annotated datasets of time-lapse microscopy images of three cell types, which are publicly available and can be used for new image analysis models. In addition, we were able to significantly increase the size of annotated data covering more variations by applying image augmentation techniques, which benefited from both conventional image augmentation techniques and a proposed random cell movements method. We expect this resource to facilitate the future application of deep learning-based models for the analysis of microscopy images.

To address the challenge of cell segmentation and tracking, we built a deep learning model, termed DeepSea, which can efficiently segment cell areas in phase-contrast microscopy images. Our segmentation model was trained on our generated dataset and achieved an IoU value of 0.90±0.2 at the IoU matching threshold of 0.5. We were able to improve on existing segmentation models by incorporating (i) an auxiliary model trained on where cell edges meet to be able to separate cells that are close to each other (ii) the addition of the residual blocks to decrease the number of parameters without sacrificing the accuracy making our model efficient (iii) progressive learning technique to improve the generalizability of our model for images with different resolution. Importantly, we were able to exploit the deep learning capabilities to automate the tracking of cells

across the time-lapse microscopy image sequences. Our DeepSea tracking model was able to track the full cell cycle trajectories with a MOTA value of 0.94±0.3 obtained from 228 cell cycles. We also showed that more frequent imaging of microscopy frames would increase the accuracy of tracking the full cell cycle by providing more information about the cell features right before cell division.

We would like to note that our dataset and models are limited to the phase contrast 2D images of three cell types. However, the researchers can train our model using their own annotated images of single cells using DeepSea software training options. A larger dataset of samples from different cell types and different imaging modalities would be useful for testing our proposed model's generalization, reliability, and robustness. In addition, in our future work, we will investigate other deep models that have recently achieved considerable advancement in object detection and tracking tasks, such as Recurrent Yolo, TrackR-CNN, JDE, RetinaNet, and CenterPoint [130-132], or merge their architecture with our current models to improve the results. To reduce the DeepSea sensitivity to the frame sampling, we will also evaluate the idea of feeding more previous frames into the tracking model, including the cell images of t, t-τ, t-2τ, and t-3τ, as one of the possible solutions.

5 cGAN-Seg: An Enhanced Cell Segmentation with Limited Training Datasets using Cycle Generative Adversarial Networks

5.1 Introduction

Generative Adversarial Networks (GANs) have gained significant attention in recent years due to their remarkable success in generating realistic images and videos [169, 170]. GANs are deep learning architectures that consist of two neural networks: a generator and a discriminator [171-173] [3-5]. The generator network is responsible for synthesizing new data, while the discriminator network attempts to distinguish between real and synthetic samples. The two networks compete in a game-like scenario until the generator produces data that is nearly indistinguishable from real data. GANs have exhibited substantial capabilities across a wide variety of applications, including image synthesis [169], video generation [174], and natural language processing [175].

Cell segmentation is a crucial step in microscopy, which involves the identification and delineation of individual cells within images. Cell segmentation is inherently complex due to the diversity and irregularity of morphological features of different cell types, such as shape and size, as well as the propensity for cells to cluster together, making highly accurate segmentation a challenging task. Deep learning methods, particularly convolutional neural networks (CNNs), have shown great success in improving cell segmentation accuracy [176-180]. The development and application of deep learning

models for cell segmentation heavily depend on the availability of a large amount of annotated training data.

While recent efforts have introduced useful large datasets of microscopy images such as LiveCELL [181], manual annotation of cell images remains a laborious and time-consuming task. Although software such as Cell-ACDC and microSAM have made significant improvements in segmentation, annotation speed, and efficiency by assisting experts in the manual labeling process [177,182], outlining individual cells in microscopic images to create 'ground truth' masks still requires considerable effort by cell biology experts. This process becomes practically infeasible when dealing with large volumes of data or when timely results are needed. Despite the progress in dataset development, the scarcity of annotated data still poses challenges to training robust models, thus slowing down progress in building cell segmentation models that work across diverse imaging scenarios and modalities.

In this study, we propose a solution to this challenge by employing a novel CycleGAN-based segmentation model, termed cGAN-Seg, that is designed to train cell segmentation models with limited annotated data. We used cGAN-Seg to generate realistic and diverse microscopy images of cells in different modalities, thereby enriching the training data for the segmentation model. We systematically compared the cGAN-Seg training approach with that of the conventional training approach on the performance of our segmentation model. Our results showed that the cGAN-Seg approach, with limited annotated cell image datasets, increased the model's ability to generalize and, thus, enhanced the performance of cell segmentation tasks.

5.2 Methods

5.2.1 Overview of designing an enhanced cycle-GAN architecture

An overview of our cGAN-Seg design is illustrated in Figure 5.1A. The cGAN-Seg architecture uses a new CycleGAN approach [183] for cell image segmentation, offering enhanced data diversity. The cGAN-Seg has several new key features: (i) Application of a novel style generation path within a 2D-UNET-based image generator [184, 185], our approach captures complex image variations from varied cell shapes to intricate structures and textures and also enhances the creation of diverse synthetic microscopy cell images (Figure 5.3). This prepares the model to tackle a wide range of image conditions, boosting its versatility and accuracy, (ii) Two adversarial PatchGan discriminators enhanced with a linear attention layer [186, 187], our model focuses on the most salient features, producing synthetic images that closely mirror real samples (Figure 5.4). This approach guarantees stable and efficient training dynamics, making the synthetic images more authentic, (iii) Implementation of a differentiable image augmentation technique [188] in our design applies identical differentiable augmentations to both real and synthetic samples (Figure 5.1A and 5.2). This reduces the risk of the discriminator memorizing exact training samples, combats overfitting, and improves the realism of synthetic images, aligning them closely with real-world scenarios, and finally (iv) Through a balanced utilization of different loss functions, including L1 and VGG-based perceptual loss functions (for the generation model) [189, 190] and Cross-Entropy and Dice losses (for the segmentation model), the model ensures the production of high-fidelity synthetic cell images while maintaining segmentation accuracy (Equations 5.6-5.17).

The CycleGAN architecture is unique in that every step in the training process encompasses two mapping paths: forward consistency and backward consistency (Equation 5.1). In the forward consistency path, the model learns to translate an image

from domain A (mask image) to domain B (cell image). Subsequently, it attempts to translate this new cell image back to domain A, aiming to reconstruct the original mask image. The backward consistency path mirrors this process, starting with domain B, translating to domain A, and then back to B. The essence of this approach lies in its ability to maintain the integrity of the original images throughout the translation and back-translation processes, ensuring that crucial information is not lost. This dual-path mechanism has the potential to introduce a greater degree of diversity to the training data. Within this architecture, the generator plays a pivotal role. It is capable of creating new synthetic cell images that might exhibit a significantly different distribution pattern from the original, real training samples. As a result, our segmentation model can train on a blend of these artificially generated images and the augmented real images. This hybrid training approach not only diversifies the data pool but also helps the model adapt to a broader spectrum of cell images. The addition of synthetic images simulates a wider array of scenarios that the model may encounter, thus improving its robustness and predictive power when faced with unfamiliar data.

$$\begin{cases} \textbf{Forward Consistency Path}: \\ \text{Real mask image} \xrightarrow{\text{Generation}} \text{Fake cell image} \xrightarrow{\text{Segmentation}} \text{Fake mask image} \\ \textbf{Backward Consistency Path}: \\ \text{Real cell image} \xrightarrow{\text{Segmentation}} \text{Fake mask image} \xrightarrow{\text{Generation}} \text{Fake cell image} \end{cases} \tag{5.1}$$

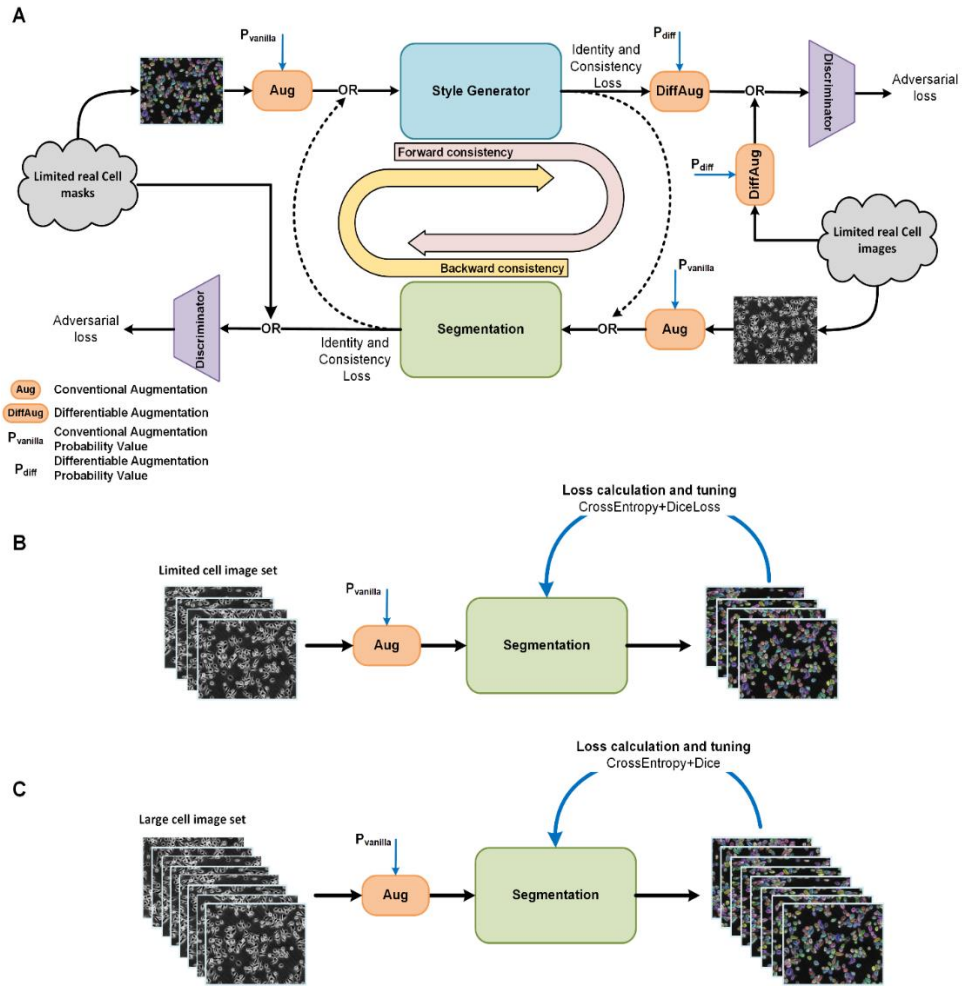

Figure 5.1: Three different cell segmentation training scenarios. A) Training cGAN-Seg on a limited dataset of cell images (e.g., 200 training samples in this paper) using a new design of CycleGAN approach, termed cGAN-Seg, that incorporates features such as style injecting, modified PatchGan discriminator, and differentiable image augmentation. B) Conventional training of the segmentation model on a limited dataset of cell images. C) Conventional training of the segmentation model on a large dataset of cell images (e.g., 1000 training samples in this paper).

5.2.2 Datasets

In our study, we utilized four distinct training datasets, each representing different modalities and cell types, to ensure the robustness and generalizability of our proposed segmentation model across a diverse range of biological contexts. They include 1) our

recently published annotated dataset of phase-contrast images of the DeepSea [180], which is a large collection of accurately annotated phase-contrast time-lapse microscopy images of three cell types of Mouse Embryonic Stem Cells, Bronchial epithelial cells, and Mouse C2C12 Muscle Progenitor Cells, 2) LiveCell dataset [181] which is a diverse collection of annotated microscopy images. It covers various cell types such as A172, BT474, Huh7, and SkBr3 across multiple experimental conditions, 3) Cell Tracking Challenge dataset [191], which is a dataset repository consisting of 2D and 3D time-lapse sequences of fluorescent images of different cell types such as PSC and U373 cells, 4) CellPose dataset [179] which is a curated collection encompassing a wide variety of annotated images from different cell types, tissues, and organisms. We used 15% of the dataset samples to test all training scenarios. To ensure a rigorous evaluation of our model, we allocated 15% of the samples from each dataset exclusively for testing across all training scenarios.

5.2.3 Augmentation functions

Image augmentation techniques play a critical role in expanding the diversity of training datasets, thereby improving model generalization and robustness [192, 193]. In the training process of our deep learning models, we applied some mostly used conventional image augmentation functions to every single cell image with the probability of p_vanilla, including random histogram equalization, random crop, random sharpness adjustment, random brightness adjustment, random contrast adjustment, random horizontal flip, random vertical flip, random saturation, adding random gaussian noise, and adding random gaussian blur as shown in Figure 5.2A. For the binary mask images, we only used the applicable random crop, horizontal flip, and vertical flip functions. The training algorithm executes a sequence of the provided augmentation functions for each cell and mask image pair with a pre-defined probability value 'p_vanilla'. In the requested

augmentation pipeline, each function is randomly chosen with a consistent probability of 50% and is also applied in a randomized sequence.

However, when it comes to training Generative Adversarial Networks (GANs), especially models like CycleGAN that learn mappings between different image domains, conventional augmentation might not be sufficient for enhancing the diversity of generated images. This is where differentiable augmentation, as proposed in [188], becomes valuable. Differentiable augmentation applies the same random augmentations to both real and fake samples in a way that is differentiable with respect to the model parameters. This approach encourages the discriminator to less memorize the exact training samples, thus causing the generator to produce more diverse images, thereby improving the overall image generation performance. Furthermore, differentiable augmentation can mitigate overfitting and improve training stability, making it particularly beneficial for GANs trained with limited data. In this project, we used five different differentiable augmentation functions: random contrast, random brightness, random cutout, random translation, and random saturation (Figure 5.2B). The decision to perform an augmentation is dictated by the probability variable 'p_diff'. In an attempt to ensure fair representation and randomness, each of the differentiable augmentation functions is executed in a randomized sequence, with each having an equal 50% probability of selection. This approach not only diversifies the images but also ensures that the model remains adaptable to any new form of data it might encounter in the future, thus improving its resilience and overall effectiveness.

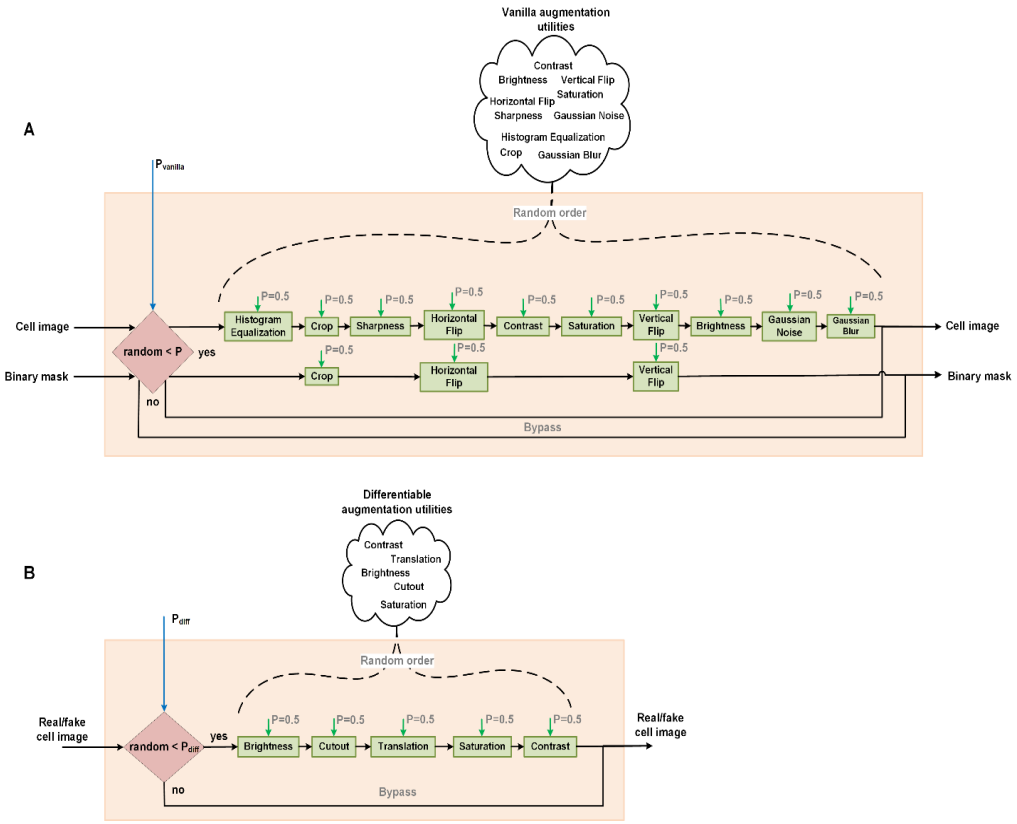

Figure 5.2: Augmentation functions used in the proposed training process. A) The conventional vanilla augmentation that was applied to both cell image and corresponding binary mask pairs aimed to increase the diversity of samples fed into the cGAN-Seg model. B) The differentiable augmentation functions used to reduce the risk of discriminator overfitting and help the generator produce more diverse synthetic images.

5.2.4 Segmentation models

For the segmentation tasks, we employed and compared our recently published DeepSea baseline architecture [180] along with two widely used segmentation models: the 2D-UNET [185] and CellPose [179] models. The 2D-UNET model has become synonymous with high-performance image segmentation across various biomedical applications. Its design is characterized by a symmetric encoder-decoder structure that efficiently captures context and enables precise localization. This architecture facilitates

the learning of rich feature representations from limited training data, making it particularly suitable for medical imaging tasks where annotated samples are scarce. The strength of 2D-UNET lies in its ability to handle a wide range of cell types and imaging conditions, thanks to its deep convolutional layers and skip connections that preserve spatial information across the network. DeepSea architecture is also an efficient scaled-down version of the 2D-UNET mode. To simplify the task, we chose not to incorporate the layer representing touching cells, as this would necessitate custom touching cell masks (alongside cell body masks) and additional loss functions. The CellPose model also represents a significant leap forward in the segmentation of complex cell images. It is designed around the concept of predicting cell 'poses'—spatial arrangements that are invariant to cell shape and size—allowing it to segment cells in a highly generalized manner. This model leverages a powerful neural network trained on a diverse dataset, enabling it to accurately segment cells across different experiments without the need for retraining. The CellPose model's robustness and adaptability stem from its novel use of flow fields, which guide the segmentation process and ensure high precision across varying biological contexts. We would like to mention that to reduce complexity and focus on the essential aspects of segmentation, we excluded CellPose's flow fields from the cGAN-Seg approach.

5.2.5 Style-based generative model

In our proposed cGAN-Seg architecture, the generator is responsible for generating synthetic cell images. It employs a 2D-UNET architecture [185], as shown in Figure 5.3. UNET is renowned for its effectiveness in biomedical image segmentation due to its unique architecture, which consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. However, we have taken this

a step further by incorporating a style decoding path into the decoder part of the UNET architecture, an idea inspired by the StyleGAN2 model [184].

This fusion of concepts from StyleGAN2 and UNET brings about the prospect of generating better synthetic images. The style decoding network is designed to control the stylistic aspects of the generated images, thereby allowing the model to create more diverse and potentially higher-quality synthetic cell images. This combination of architectures seeks to maximize the strengths of both models - the segmentation prowess of UNET and the sophisticated generative capacity of StyleGAN2. This integration could potentially yield a more powerful generator model for synthetic cell image creation, thereby enhancing the overall performance of our cGAN-Seg model.

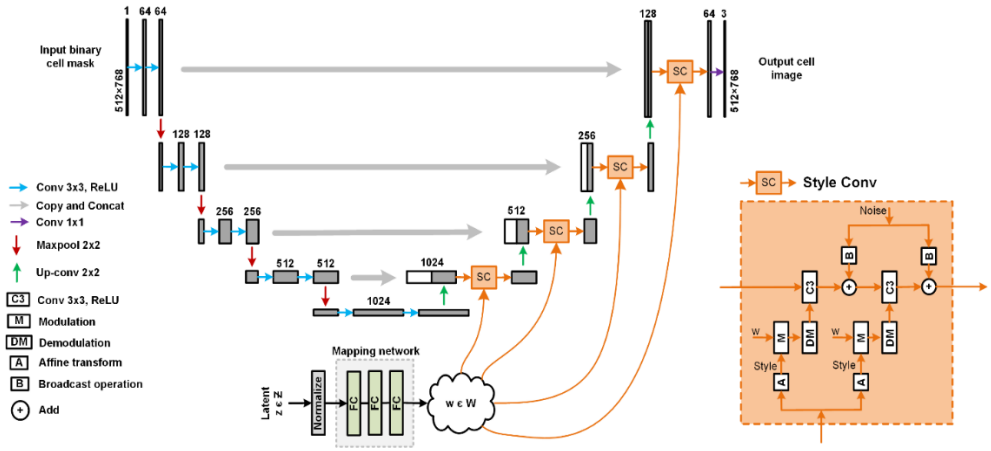

Figure 5.3: Our proposed generative architecture. It employs the 2D-UNET architecture with a style decoding network to create more diverse and potentially higher-quality synthetic cell images.

5.2.6 Discriminators

In our proposed method, we applied a modified version of the PatchGAN baseline architecture [186] for our discriminators, integrating a layer of residual linear attention, as

shown in Figure 5.4. The PatchGAN architecture, known for its effectiveness in examining both global and local image features, has demonstrated impressive performance in diverse GAN applications. The architecture operates by creating multiple 'paths' with different receptive field sizes, enabling the model to scrutinize image details at various scales.

However, we sought to improve the discriminator's ability to focus on critical features by incorporating an additional layer of residual linear attention [187]. This is an approach to attention mechanisms that makes use of a linear combination of input features and learned attention maps, thereby enabling the model to weigh different regions of the input differently. As a result, the model can focus on more critical parts of the image, thereby enhancing its ability to discriminate real images from synthetic ones accurately. By modifying the PatchGAN discriminator, we aimed to improve the model's focus on salient image features, thus boosting its ability to accurately distinguish between real and generated images.

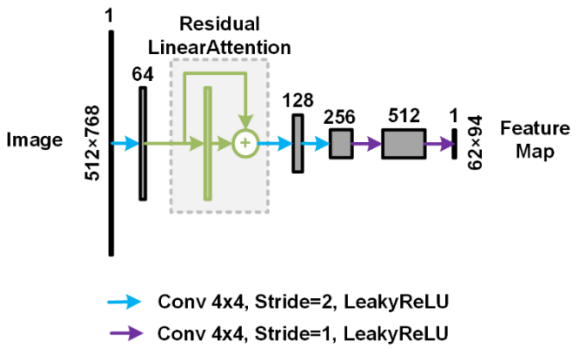

Figure 5.4: A modified version of the PatchGan discriminator used in our proposed cGAN-Seg. We integrated a layer of residual linear attention to improve the discriminator performance.

5.2.7 Evaluation metrics

In the testing phase, we leveraged the Intersection over Union (IoU) metric, also known as the Jaccard Index, which ranges from 0 to 1, to evaluate the alignment between the segmentation model's predictions and the manually annotated ground truth masks [194]. For each test image, we designated each detected cell body as a True Positive (TP) prediction if its IoU index exceeded a predetermined threshold value, indicating a valid match to the ground truth. Conversely, any ground truth cell body masks that failed to find a valid match were classified as False Negatives (FN), and any predictions lacking corresponding ground truth masks were labeled as False Positives (FP), representing non-cell entities. Subsequently, we calculated the Precision, Recall, and F-score for each image in the test set using Equations (5.2-5.4).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{5.3}$$

$$\text{Fscore} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{5.4}$$

We also evaluated our segmentation models using Dice Score. The Dice score, also known as the Dice Similarity Coefficient, is a common metric used in the field of image segmentation, including cell image segmentation, to measure the similarity between two samples and can be measured by Equation (5.5), X represents the set of pixels in the predicted segmentation and Y represents the set of pixels in the ground truth segmentation.

$$\text{Dice Score} = 2 \times \frac{|\text{X} \cap \text{Y}|}{|\text{X}| + |\text{Y}|} \tag{5.5}$$

Frechet Inception Distance (FID) is also a widely used metric to evaluate the quality of images generated by GAN models [195]. It quantifies the dissimilarity between the

distributions of generated and real images in the feature space of a pre-trained Inception network. Lower FID scores represent higher-quality synthetic images that more closely resemble the distribution of real images. By assessing differences in both mean and covariance of the features, FID provides a more comprehensive evaluation of image quality and diversity, making it a good choice for evaluating the performance of our modified generator.

5.2.8 Loss functions

In the training phase of our cGAN-Seg model, we employed a series of loss functions to effectively optimize the performance of both the generator and the segmentation model. These loss functions are specifically tailored to address the unique challenges presented by the task of generating high-quality synthetic images and accurately segmenting cell structures.

Two fundamental loss functions utilized in our model are identity loss and reconstruction loss. Identity loss ensures that an image translated to its own domain remains unchanged, which encourages the generator to preserve color and texture composition between the input and output [183]. Reconstruction loss, on the other hand, is used to maintain cycle consistency, ensuring that an image translated from one domain to another can be accurately translated back to its original form. These losses are critical to ensure that the model not only learns the correct mappings between the domains but also produces images that are consistent with the original data distribution.

For the generator, we utilized the VGG perceptual feature loss function [190, 196] for both identity and reconstruction loss (Equations 5.8 and 5.10). The VGG loss function is a high-level feature extraction loss that helps preserve the perceptual and semantic understanding of the images. It is a concept based on a deep convolutional neural network (CNN), like VGG, that has been pre-trained on a large dataset for an image classification task.

For the segmentation model, we combined Cross-Entropy (CE) loss and Dice loss for both identity and reconstruction loss (Equations 5.9 and 5.11). Cross-Entropy loss is a popular choice for multi-class classification problems, calculating the dissimilarity between the predicted probability distribution and the ground truth distribution. The Dice loss, on the other hand, is specifically designed for handling imbalanced datasets and is extensively used in medical image segmentation tasks due to its efficiency in dealing with small objects and imbalanced classes. By using these two loss functions in tandem, we enhance the performance of our segmentation model, ensuring it can effectively handle the challenges of cell image segmentation.

The discriminators in our model were optimized using the Mean Squared Error (MSE) loss as an adversarial loss. This loss function encourages the discriminators to distinguish between real and fake images by minimizing the average squared differences between the predicted and actual values.

Each of these loss functions is assigned a specific weight in order to balance their contributions during the optimization process (Equations 5.16 and 5.17). By integrating these diverse loss functions and carefully selecting their weights, we can effectively train our cGAN-Seg model, ensuring both the production of diverse, high-quality synthetic cell images and the accurate segmentation of cell structures.

$$\mathrm{D1\_G\_L} = \mathrm{MSE}(1, \mathrm{D1}(\mathrm{DiffAug}(\mathrm{Gen}(\mathrm{real\_mask})))) \tag{5.6}$$

$$\mathrm{D2\_S\_L} = \mathrm{MSE}(1, \mathrm{D2}(\mathrm{Seg}(\mathrm{real\_img}))) \tag{5.7}$$

$$\mathrm{Rec\_G\_L} = \mathrm{L1}(\mathrm{real\_img}, \mathrm{Gen}(\mathrm{Seg}(\mathrm{real\_img}))) + \mathrm{VGG}(\mathrm{real\_img}, \mathrm{Gen}(\mathrm{Seg}(\mathrm{real\_img}))) \tag{5.8}$$

$$\mathrm{Rec\_S\_L} = \mathrm{CE}(\mathrm{real\_mask}, \mathrm{Seg}(\mathrm{Gen}(\mathrm{real\_mask}))) + \mathrm{Dice}\left(\mathrm{real\_mask}, \mathrm{Seg}\left(\mathrm{Gen}(\mathrm{real\_mask})\right)\right) \tag{5.9}$$

$$\text{Id\_G\_L} = \text{L1}(\text{real\_img}, \text{Gen}(\text{real\_mask})) + \text{VGG}(\text{real\_img}, \text{Gen}(\text{real\_mask})) \quad (5.10)$$

$$\text{Id\_S\_L} = \text{CE}(\text{real\_mask}, \text{Seg}(\text{real\_img})) + \text{Dice}(\text{real\_mask}, \text{Seg}(\text{real\_img})) \quad (5.11)$$

$$\text{Real\_D1\_L} = \text{MSE}(1, \text{D1}(\text{DiffAug}(\text{real\_img}))) \quad (5.12)$$

$$\text{Real\_D2\_L} = \text{MSE}(1, \text{D2}(\text{real\_mask})) \quad (5.13)$$

$$\text{Fake\_D1\_L} = \text{MSE}(0, \text{D1}(\text{DiffAug}(\text{Gen}(\text{real\_mask})))) \quad (5.14)$$

$$\text{Fake\_D2\_L} = \text{MSE}(0, \text{D2}(\text{Seg}(\text{real\_img}))) \quad (5.15)$$

$$\text{Total\_Gen\_loss} = \text{D1\_G\_L} + \text{D2\_S\_L} + 100 \times \text{Rec\_G\_L} + 100 \times \text{Rec\_S\_L} + 50 \times \text{Id\_G\_L} + 50 \times \text{Id\_S\_L} \quad (5.16)$$

$$\text{Total\_D\_loss} = 0.5 \times \text{Real\_D1\_L} + 0.5 \times \text{Real\_D2\_L} + 0.5 \times \text{Fake\_D1\_L} + 0.5 \times \text{Fake\_D2\_L} \quad (5.17)$$

5.2.9 Code and data availability

The Python scripts encompassing the methodologies we developed are publicly accessible for download at our GitHub repository: https://github.com/abzargar/cGAN-Seg. Additionally, the image dataset utilized in our study is available via a link on the repository's page, facilitating easy access for replication and further research efforts.

5.3 Results

5.3.1 Model's performance evaluation

To test our proposed cGAN-Seg approach, we designed and ran three distinct segmentation training scenarios: 1) Employing cGAN-Seg to train segmentation models on a selected small subset of a dataset, followed by evaluation on the selected test set (Figure 5.1A); 2) Direct and conventional training of segmentation models on the same small subset without cGAN-Seg, with subsequent testing on the same test set (Figure 5.1B); 3)

Direct and conventional training of segmentation models on the full dataset of the same cell type and assessing its performance on the same test set (Figure 5.1C). To ensure the robustness and generalizability of our proposed approach across a diverse range of biological contexts, we utilized samples from four different image datasets, each featuring different modalities and cell types. These datasets include DeepSea [180], the LiveCell dataset [181], the Cell Tracking Challenge dataset [191], and the CellPose dataset [179]. For the segmentation tasks, we used the DeepSea baseline architecture [180], as well as two of the most widely utilized baselines in the field: the 2D-UNET [185] and the CellPose models [179]; The 2D-UNET is known for its high performance and suitability for biomedical imaging due to its encoder-decoder architecture, the DeepSea architecture is a streamlined adaptation of 2D-UNET optimized for simplicity without compromising on efficiency; and CellPose uses a versatile neural network that is trained on diverse cell types, allowing for generalization across various cell types.

In the testing phase, our evaluation focused on the accuracy of the segmentation models by measuring the Intersection Over Union (IOU) between predicted and true cell masks. Cells detected with an IOU at or above our threshold were labeled as the correct detection. To provide a holistic view of model performance, we calculated precision, recall, and the F-score. These metrics inherently account for the impact of false positives (incorrectly predicted cells) and false negatives (missed actual cells), offering a balanced measure of our models' effectiveness across varying IOU thresholds (Equations 5.2-5.4). In all experiments and score reports, we applied the five-fold cross-validation technique, aiming to provide a more robust assessment of the segmentation model's performance by reducing the impact of random variations in the training and validation data splits and ensuring that the model's performance is not overly influenced by a specific subset of the data.

Figure 5.5 presents the average f-score of three segmentation models trained using the cGAN-Seg method on a small dataset of 200 samples. This performance is compared with that achieved by conventional training methods (Figures 5.1B and 5.1C), using both the same small dataset and a larger dataset of 1000 samples. It focuses on two distinct cell types from the DeepSea dataset for which sufficiently large, annotated image collections were available. In all test experiments, cGAN-Seg significantly improved the performance of three segmentation models of 2D-UNET, CellPose, and DeepSea almost at all IOU threshold values compared to when we trained the segmentation model with the same limited dataset. This enhancement is particularly evident in stem cell samples, which pose a greater challenge due to the higher diversity and complexity of cell images. Importantly, the segmentation scores achieved using the cGAN-Seg method with a limited dataset closely match, or even surpass, those of segmentation models trained with traditional approaches on large datasets. This outcome underscores the cGAN-Seg architecture's style generator's capability to produce a diverse enough array of samples during training. This diversity effectively compensates for the limited data available to the segmentation model, showcasing the method's efficiency in optimizing performance despite data constraints.

Next, we extended our analysis to two other publicly available datasets or microscopy images, LiveCell and Cell Tracking Challenge, to test the applicability of cGAN-Seg for a diverse set of cell types. Table 5.1 summarizes an f-score comparison between the conventional segmentation and our proposed cGAN-Seg training method for different cell types in the LiveCell and Cell Tracking Challenge datasets. In addition, we measured the average recall, precision, and dice score metrics (Tables 5.2-5.4) for the 2D-UNET and CellPose models. The results demonstrate that the cGAN-Seg model improves the segmentation model performance in every case. This consistent enhancement across

different cell types and experimental conditions highlights the model's robustness and its adaptability to the inherent variations present in biological imaging data, particularly where data scarcity often limits the efficacy of conventional approaches.

Next, we sought to test the impact of the modifications we introduced in the CycleGAN-based approach on the generation of synthetic images of cells by comparing synthetic cell images generated by three different generator configurations: 1) UNET generator with L1 loss, 2) StyleUNET generator with L1 loss, and 3) StyleUNET generator with L1+VGG loss (Figure 5.6) and used the DeepSea model for the segmentation part. Synthetic images generated by the UNET model with L1 had an average Frechet Inception Distance (FID) score of 98, indicating a high dissimilarity between the synthetic and real cell images. The FID score dropped to 43 when we applied StyleUNET with L1 loss, resulting in increased similarities between synthetic images and real images. This improvement primarily stems from the enhanced capability of StyleUNET to capture and generate variations in style. Lastly, when we applied StyleUNET with the VGG perceptual loss, the synthetic images achieved a remarkably enhanced FID score of 23, reflecting a substantial increase in similarity to real cell images. The overall results of these modifications are synthetic images with detailed representations of features at subcellular levels. The enhanced similarity of synthetic images and lower FID score signifies the superiority of the perceptual loss function in preserving high-level details and morphological nuances, thereby leading to more realistic synthetic images.

To demonstrate the extensibility of our proposed approach, we trained the cGAN-Seg model across various imaging modalities, diverse cell types, and subcellular organelles (from DeepSea and CellPose datasets). Subsequently, the trained StyleUNET generator was employed to produce synthetic images that span these diverse conditions. While the limited number of available annotated images posed a challenge for a subset of

samples, our SyleUNET generation model still generated realistic synthetic images similar to the real domain images and relatively overall low FID scores (Figure 5.7). We anticipate that with more extensive training data, the model's capacity for generating high-quality images would be significantly enhanced.

Most of our training data consisted of low and mid-density cell images. We were curious to test if our generator could extrapolate its knowledge by generating new synthetic high-density images of cells that were not seen during training. This is particularly useful as manual annotation of high-density images of cells can be very time-consuming and error-prone. To accomplish this, we designed an algorithm for generating synthetic high-density and colony-like cell masks (a relatively easy task) as input for the generator (in the test phase). As shown in Figure 5.8, our approach confirms the ability to extrapolate knowledge from low and mid-density cell images, creating annotated images across any density level and magnification. This also includes the generation of colony-like cell formations (Example 3 in Figure 5.8) extending beyond the variations present in the original dataset. Such a capacity enables us to create synthetic cell images mimicking a broad range of real-world scenarios. The capacity of our model to extrapolate learned knowledge to unseen scenarios can provide a powerful tool to generalize this approach, aiming to develop segmentation models for a variety of cellular imaging modalities.

To further assess the capability of the StyleUNET generation model in enhancing the segmentation of more challenging, high-density cell images, we used our synthetic high-density cell images (showcased in Figure 5.8) to conventionally train a segmentation model. To validate the model evaluation process, we categorized the test set images into two groups based on their complexity: 'easy' samples, which include isolated cells or cells in non-touching colonies, and 'hard' samples, characterized by cells in close contact or within touching colonies. In this experiment, we employed the DeepSea model for the

segmentation and focused on the DeepSea stem cell test set, where we found enough easy and hard samples for this specific analysis. As presented in Table 5.5, integrating hard, colony-like, and high-density synthetic images into the training process notably enhances segmentation performance on 'hard' samples to a greater extent compared to the 'easy' samples, underscoring the effectiveness of StyleUNET generation model outputs in tackling complex segmentation scenarios.

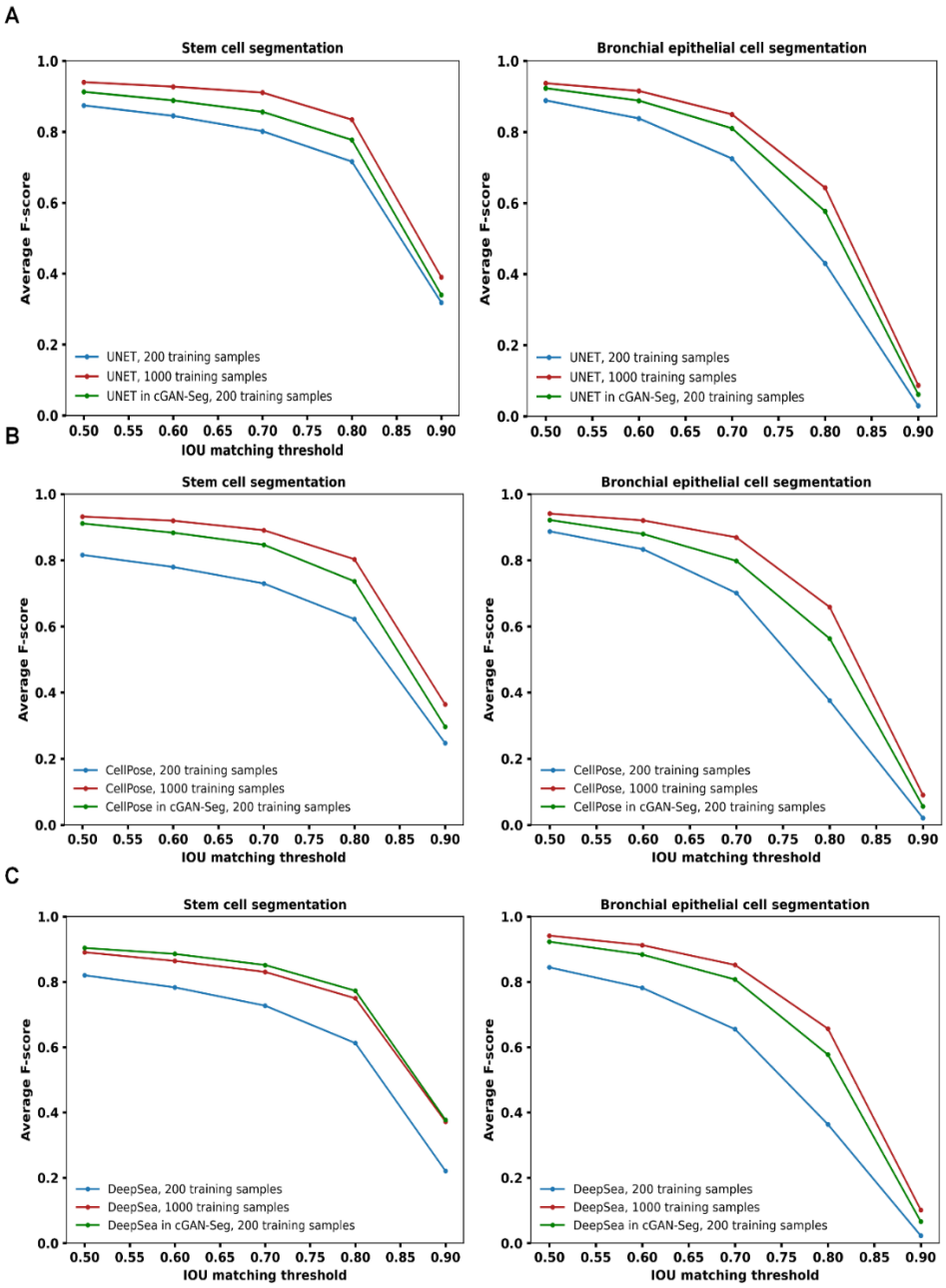

Figure 5.5: F-Score performance across varied IoU thresholds for A) UNET (top row), B) CellPose (middle row), and C) DeepSea (bottom row) segmentation models using cGAN-Seg training vs. conventional training on datasets of 200 and 1000 samples.

Table 5.1: F-Score comparison for cGAN-Seg and conventional segmentation. Comparative analysis of average f-scores at 0.5 IoU threshold: cGAN-Seg segmentation vs. conventional segmentation using limited datasets (200 training samples) across diverse cell types.

	A172 [181]	BT474 [181]	Huh7 [181]	SkBr3 [181]	C2C12 Muscle [180]	PSC [191]	U373 [191]
CellPose	0.51	0.53	0.47	0.89	0.79	0.90	0.95
2D-UNET	0.52	0.53	0.49	0.90	0.79	0.92	0.96
cGAN-Seg Seg: CellPose	0.60	0.56	0.55	0.91	0.84	0.92	0.98
cGAN-Seg Seg: 2D-UNET	0.60	0.55	0.59	0.92	0.83	0.95	0.98

Table 5.2: Precision comparison for cGAN-Seg and conventional segmentation. Comparative analysis of average precision at 0.5 IoU threshold: cGAN-Seg vs. conventional segmentation using limited datasets (200 training samples) across diverse cell types.

	A172 [181]	BT474 [181]	Huh7 [181]	SkBr3 [181]	C2C12 Muscle [180]	PSC [191]	U373 [191]
CellPose	0.63	0.56	0.57	0.89	0.82	0.90	0.91
2D-UNET	0.63	0.57	0.60	0.90	0.83	0.91	0.95
cGAN-Seg Seg: CellPose	0.64	0.57	0.59	0.90	0.87	0.90	0.96
cGAN-Seg Seg: 2D-UNET	0.63	0.58	0.63	0.91	0.87	0.94	0.97

Table 5.3: Recall comparison for cGAN-Seg and conventional segmentation. Comparative analysis of average recall at 0.5 IoU threshold: cGAN-Seg vs. conventional segmentation using limited datasets (200 training samples) across diverse cell types.

	A172 [181]	BT474 [181]	Huh7 [181]	SkBr3 [181]	C2C12 Muscle [180]	PSC [191]	U373 [191]
CellPose	0.44	0.51	0.41	0.89	0.76	0.90	0.98
2D-UNET	0.45	0.50	0.42	0.89	0.75	0.92	0.96
cGAN-Seg Seg: CellPose	0.57	0.56	0.52	0.90	0.82	0.93	0.99
cGAN-Seg Seg: 2D-UNET	0.57	0.53	0.56	0.91	0.80	0.95	0.98

Table 5.4: Dice score comparison for cGAN-Seg and conventional segmentation. Comparative analysis of average dice score at 0.5 IoU threshold: cGAN-Seg vs. conventional segmentation using limited datasets (200 training samples) across diverse cell types.

	A172 [181]	BT474 [181]	Huh7 [181]	SkBr3 [181]	C2C12 Muscle [180]	PSC [191]	U373 [191]
CellPose	0.87	0.80	0.87	0.90	0.87	0.93	0.97
2D-UNET	0.87	0.80	0.87	0.90	0.87	0.94	0.98
cGAN-Seg Seg: CellPose	0.85	0.78	0.83	0.89	0.88	0.94	0.98
cGAN-Seg Seg: 2D-UNET	0.84	0.78	0.83	0.90	0.88	0.95	0.98

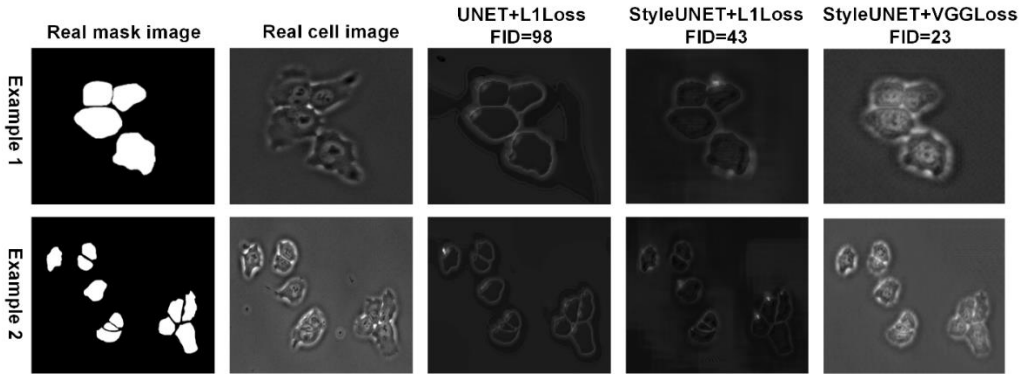

Figure 5.6: Two examples of comparing the effect of the style injecting technique and Vgg perceptual feature loss function on cGAN-Seg performance to generate images of DeepSea embryonic stem cells.

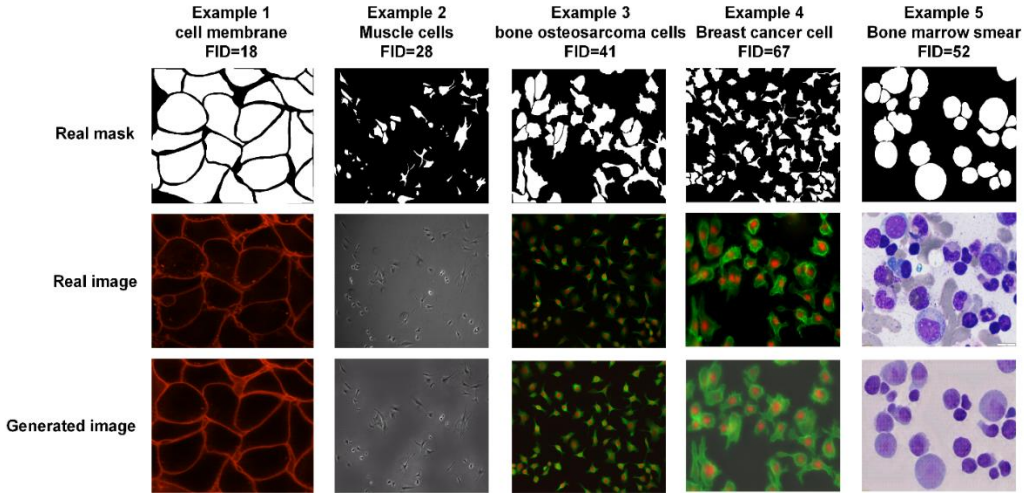

Figure 5.7: cGAN-Seg can generate synthetic images across multiple imaging modalities, cell types, and subcellular organelles (from DeepSea and CellPose datasets) similar to real images, as relatively low FID scores show.

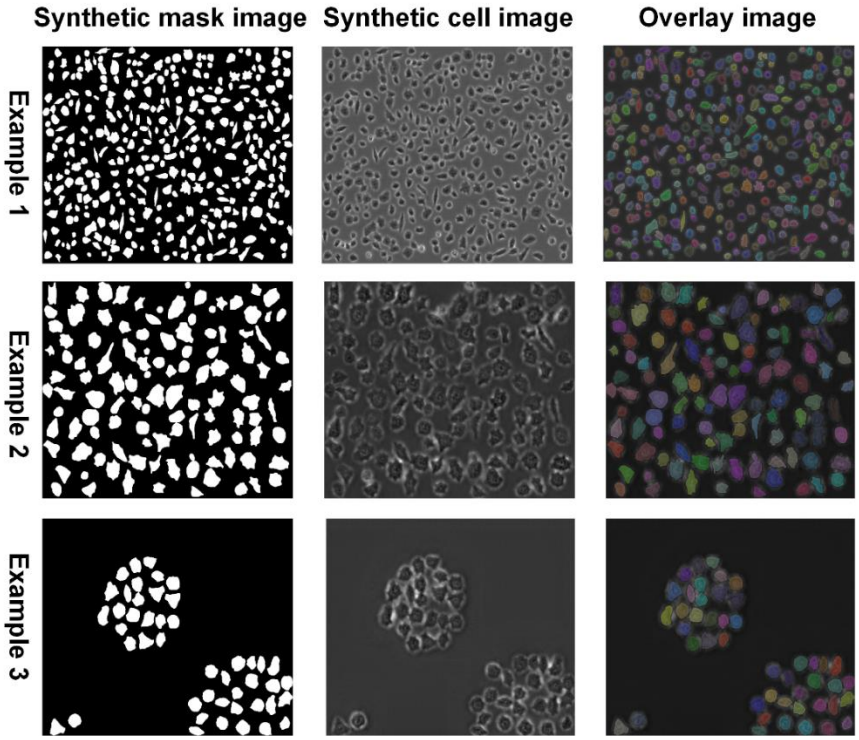

Figure 5.8: Three examples of producing colony-like and high-density cell images using synthetic high-density mask images as input for cGAN-Seg.

Table 5.5: Average f-score comparison before and after adding synthetic high-density cell images (generated by the cGAN-Seg generation model) into the conventional training process of the DeepSea segmentation model.

	200 training images	200 training images+700 synthetic images
Easy samples	0.86	0.92
Hard samples	0.77	0.86

5.4 Conclusion

A large and diverse annotated dataset of images is key to the successful development of deep learning models that can perform across a variety of real-world images. Currently, a harmonized large and diverse dataset of microscopy images is not available to train new deep-learning models because the annotation of microscopy images

is a tedious and time-consuming task. Our study provides a solution to this problem by proposing a novel method for training cell segmentation models using a CycleGAN approach that we termed cGAN-Seg to address the critical issue of limited annotated data in cellular imaging. cGAN-Seg harnesses the potential of GANs to generate a diverse set of synthetic realistic cell images, enhance the diversity of the available training datasets without manual annotation, and improve the overall performance of the segmentation models with a limited annotated dataset. Importantly, we showed that cGAN-Seg allows for the extrapolation of knowledge by model by generating synthetic images that the model has not been exposed to during the training.

We made several modifications to the original CycleGAN architecture to build the cGAN-Seg model and apply the microscopy images. First, a style generation path was integrated into the synthetic image generator to boost variation in synthetic images. Second, a linear attention mechanism was incorporated into the PatchGAN discriminator-based architecture to enhance its differentiation capabilities and synthetic image quality. Third, differentiable image augmentation was introduced during the training phase to further diversify image generation and reduce the risk of overfitting. Fourth, instead of the L1 loss function conventionally used in the CycleGAN, we employed a combination of Cross-Entropy (CE) and Dice losses for the segmentation, improving the handling of multi-class classification and imbalanced datasets. Finally, as a critical modification, we replaced the L1 loss function in the generator with a VGG perceptual loss function to promote the retention of more high-level features and nuances in the generated synthetic images, leading to enhanced similarity between real images of cells and synthetic images. These enhancements collectively improved the diversity and quality of synthetic cell images, resulting in a more diverse and generalized segmentation model trained with various microscopy imaging styles and conditions.

Our experimental results show that the proposed cGAN-Seg approach provides a straightforward solution for the paucity of annotated microscopy data for training deep-leaning models. Experimental results showed that the performance of segmentation models trained using our cGAN-Seg method improved the segmentation scores across different cell types of different datasets. Notably, this enhancement was observed irrespective of the scarcity of annotated cell image datasets, illustrating the potential of our approach in effectively addressing this prevalent issue in biomedical imaging. Implementation of Style injecting in our UNET generator significantly improved the quality of synthetic images, reflected by an FID score reduction from 98 to 43. A further enhancement was achieved by adding VGG perceptual loss to the conventional L1 loss function, resulting in an FID score of 23. We also validated the model's versatility across multiple imaging modalities, cell types, and subcellular organelles. Besides, we further validated the trained cGAN-Seg ability to extrapolate knowledge from low and mid-density cell images, creating annotated images across different density levels and magnifications, even those absent in the original training dataset. We extended the utility of the cGAN-Seg model by incorporating the generated synthetic high-density cell images into the segmentation finetuning process. This addition enhanced the model's segmentation capabilities, especially for complex, densely populated cell structures.

In our experiments, we noticed that the impact of the cGAN-Seg training approach varied between different cell-type images. The dataset comprising less complexity and lower diversity showed a lower improvement compared to the more complex dataset. This can be attributed to the inherent simplicity of the data, which likely enabled the model to learn necessary patterns without the need for additional augmented examples. Conversely, for complex datasets with more inherent variability, cGAN-Seg proved more beneficial by providing diverse synthetic image examples during the training, thus enhancing the model's

ability to generalize. We believe that the exact dynamics depend on the specific dataset, model, and augmentation techniques used, underscoring the need for task-specific experimentation and validation.

Our CycleGAN-based method opens up new possibilities for training deep-learning models for microscopy applications by offering a novel solution to the challenge of limited annotated cell image datasets. This study illustrates how generative deep learning methods like GANs can be utilized to address data limitations in microscopy, thereby pushing the boundaries of what is possible in the field of biomedical imaging. It is important to mention that while our approach has shown promising results, there is room for further improvement and experimentation, including exploring different GAN architectures and further refinement of the augmentation techniques. The generated synthetic images can also be made more diverse and realistic through additional modifications in the GAN training process.

6 tGAN: A GAN-based Super-Resolution Video-to-Video Time-Lapse Microscopy Generative Model

6.1 Introduction

Recent advancements in deep learning, particularly in generative adversarial networks (GANs) [197], have revolutionized various fields, offering unprecedented capabilities in data generation and analysis [198, 199]. GAN models, known for their ability to generate highly realistic synthetic data, have found applications across a wide range of disciplines, from art creation to medical imaging [200-202]. Their importance in biomedical imaging cannot be overstated, as they provide innovative solutions to simulate realistic biological data, which is essential for training and testing analytical models [203]. Among these applications, the generation of synthetic cell images from time-lapse microscopy using deep learning models, especially GANs, stands out as a particularly promising area [204]. These models offer new avenues for research in cell biology, enabling the creation of detailed and accurate representations of cellular processes.

Time-lapse microscopy, a crucial tool in cell biology, captures the dynamic behaviors of cells over time. However, the analysis of these images poses significant challenges, particularly in tracking and interpreting complex cellular dynamics [205]. A key hurdle in this process is the lack of annotated datasets, which are vital for training deep learning models to accurately perform various tasks related to cell analysis, such as segmentation, tracking, and behavior prediction [206]. Annotated time-lapse microscopy

images are indispensable for understanding cellular mechanisms and responses, yet creating such datasets is often labor-intensive and requires expert knowledge.

In response to these challenges, this paper introduces a video-to-video generative approach designed to generate annotated time-lapse microscopy images. Leveraging the capabilities of deep learning, our model transforms binary mask image sequences into corresponding high-resolution, synthetic images of cells. This approach not only enhances the quality and utility of microscopic images but also opens up possibilities for further computational analysis and interpretation in biomedical research. The model offers a significant advancement in the automated generation of annotated datasets, facilitating more efficient and accurate analysis of cellular dynamics.

6.2 Methods

6.2.1 Overview of designing a novel Gan-based super-resolution video generator

The architectural design of our video-to-video generative model is central to its performance and is characterized by its two-part structure, encompassing both video-to-video low-resolution and image-to-image super-resolution models (Figure 6.1A). While the low-resolution model adequately captures essential dynamic structures for many applications, integrating a separate super-resolution model further enhances the quality of synthetic images. This two-part structure is designed to optimize computational efficiency and precision. Since the low-resolution model training process involves multiple components such as discriminators, flow networks, and the video-to-video generator (Figure 6.1B), integrating all these elements at a high resolution would significantly increase computational costs and could potentially degrade the performance and training efficiency of the generator. Therefore, we initially trained the video-to-video generator in low resolution, achieving higher accuracy with less computational overhead. Subsequently,

we trained a GAN-based super-resolution model (Figure 6.1C) to refine the video-to-video model outputs to high resolution, enhancing the quality and adding necessary details crucial for applications requiring high fidelity. This sequential approach not only ensures detailed, high-quality outputs but also maintains manageable computational demands, facilitating more efficient processing and superior performance in scenarios that require detailed, high-resolution images.

The low-resolution video-to-video generator, inspired by 2D-UNET architecture [207], encodes multiple inputs, including current n (for example, two) consecutive mask frames, a previously generated cell frame (at t-T), and a consistent reference background image, ensuring context-aware and consistent background generation (Figure 6.1B and Figure 6.2). One of the distinctive features of this phase is the inclusion of a reference background image. Incorporating a reference background image in our low-resolution GAN model enhances contextual accuracy and ensures consistency in background features. It also increases the variability of background visual features in synthetic samples, accommodating various real-world scenarios like background noises, debris, or optical artifacts common in microscopy. This approach enriches the realism and detail of the generated images, which is crucial for precise analysis in applications such as time-lapse microscopy, where accurate background representation is vital. The integration of attention layers allows for adaptively integrating visual features from three model inputs. The innovative inclusion of style and noise injection into the decoding path, which was introduced and validated in our prior research [208], adds variability and realism to the generated images.

Transitioning to the super-resolution phase, our image-to-image generator employs an enhanced UNET-based architecture, incorporating style and noise injection for refining textural details and aesthetic elements, adding finer details and thus resulting in

higher quality and more detailed images (Figure 6.1C and Figure 6.3). While, in our experiments, we targeted a high resolution of 512x768 for the super-resolution model output, our model is adaptable and can be easily configured to achieve higher resolutions by adjusting its parameters, thus accommodating a diverse range of research requirements. The discriminators, designed with a PatchGAN architecture [209] and enhanced with a linear attention layer [210], effectively distinguish fine details in images (Figure 6.1B and Figure 6.4). This enhances the model's accuracy in differentiating between real and synthetic images. In the training process of the video-to-video generator, alongside the GAN and discriminator models, we concurrently trained and used a FlowNET (Figure 6.1B and Figure 6.5) designed to calculate and integrate flow loss into the training regime. This FlowNET plays a crucial role in determining the optical flow loss [211], comparing motion between consecutive frames in both real and generated sequences. This is essential for preserving the dynamic nature of cell movements in time-lapse microscopy. The optical flow loss computed by our FlowNet ensures that the temporal coherence and motion consistency of generated images align closely with actual microscopy sequences. However, it is not incorporated in the super-resolution model, which concentrates on image-to-image translation rather than the generation of video sequences, thereby making flow consistency less relevant in that context.

Besides, the chosen loss functions, including temporal consistency and perceptual losses, ensure temporal coherence and visual similarity between generated and real images (Equations 6.1-6.3). Throughout the training process, we also employed robust augmentation strategies, including video-level augmentations for general model training, as well as video-level differentiable augmentations [212] for the discriminators (Figure 6.6). These augmentations enhance the model's robustness and prevent overfitting, especially when the training dataset size is limited, as proved in our previous research [208].

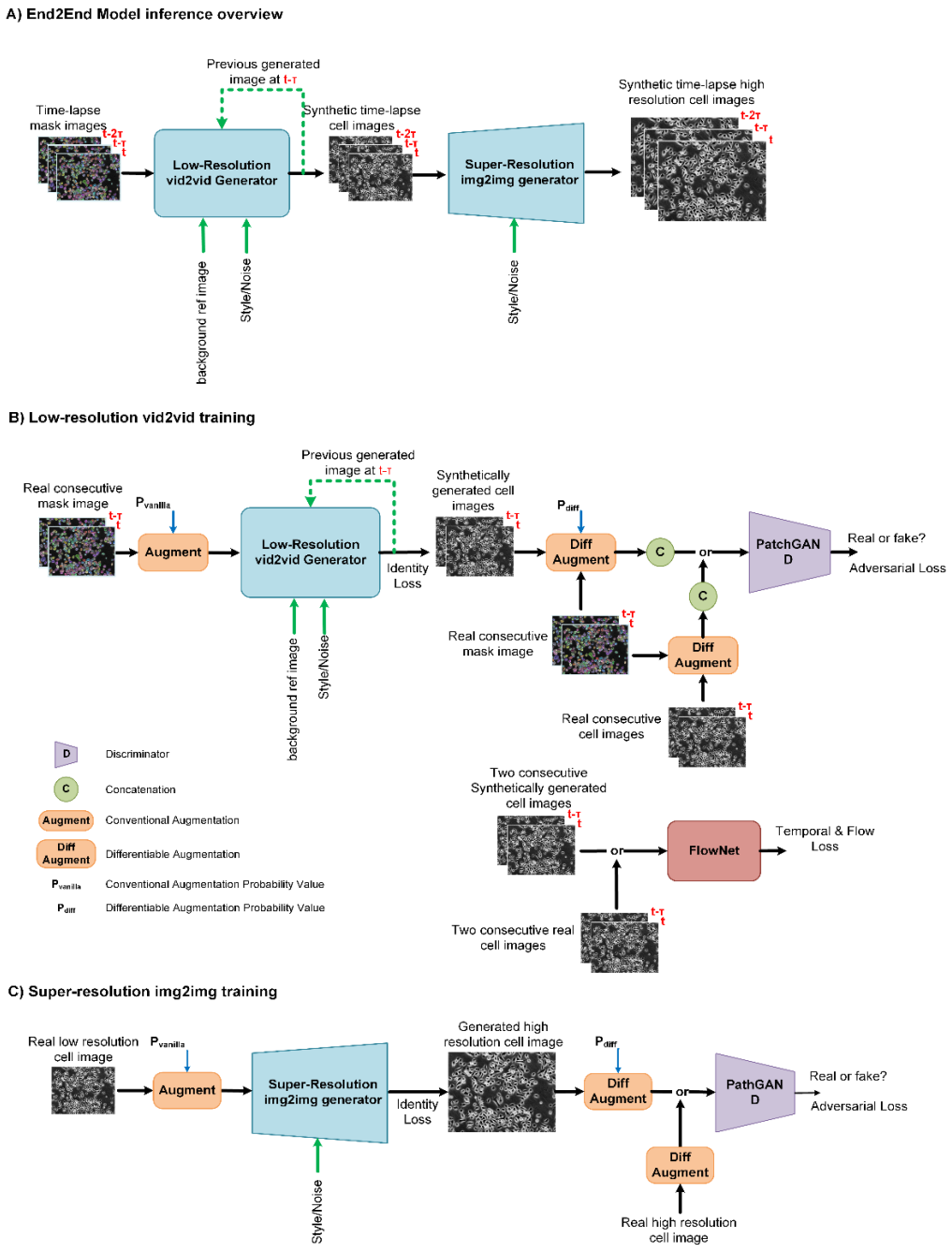

Figure 6.1: Overview of the proposed GAN-based generator. A) Illustration of the two-part inference architecture: This includes both the low-resolution video-to-video and super-resolution image-to-image models, where the initial phase generates low-resolution time-lapse microscopy image sequences to capture essential cellular dynamics, followed by the super-resolution phase that refines these sequences into more detailed and high-quality images. B) Low-resolution video-to-video training process: Showcasing the sequence-

based approach and the integration of various inputs and deep learning components. C) Super-resolution image-to-image training process: Highlighting the approach and models used for refining textural details and enhancing aesthetic elements, thereby producing scientifically accurate and visually high-resolution images.

6.2.2 Datasets

In our study, we used two distinct training datasets from time-lapse microscopy, each representing different cell types, to ensure the robustness and generalizability of our proposed generative model across various biological contexts. These include 1) our recently published annotated dataset of phase-contrast images from the DeepSea collection [213], which comprises a large set of accurately annotated phase-contrast time-lapse microscopy images of three cell types of Mouse Embryonic Stem Cells, Bronchial Epithelial Cells, and Mouse C2C12 Muscle Progenitor Cells; and 2) the Cell Tracking Challenge dataset [214], a repository of 2D and 3D time-lapse sequences of fluorescent images featuring different cell types such as PSC and U373 cells.

6.2.3 Low-resolution video-to-video generative model overview

As illustrated in Figures 6.1B and 6.2, our low-resolution model is a sequence-based model inspired by the 2D-UNET architecture, known for its efficiency in image segmentation tasks [207]. This model is uniquely designed to process three types of input: the current consecutive mask images (e.g., two consecutive mask images), the previous generated synthetic cell image, and a consistent background image that serves as a reference for background visual features. The mask sequences act as the driving force, dictating the formation of the synthetic cell images, which are then superimposed onto the reference background. This approach allows for a more realistic and context-aware generation of cell images, which is essential for accurate subsequent analysis and ensures the consistency of the background visual features as well.

Central to our model's architecture is the incorporation of specialized attention layers. These layers are strategically placed to merge features from three encoded inputs adaptively at different down-sampling stages. This process employs an attention mechanism, directing the model's focus to the most pertinent features across the different inputs, thereby enhancing the detail and relevance of the generated images. In addition to structured feature integration, our model innovatively employs style and noise injection techniques in its decoding pathway, inspired by advancements in neural style transfer [215]. As proved in our previous research [208], this approach introduces a layer of variability and texture to the synthetic images, elevating their realism and authenticity. The style and noise elements are carefully modulated to complement the intrinsic features of the cell images, ensuring that the synthetic outputs are not only high in resolution but also rich in biological detail. Its ability to process and integrate multiple input types, coupled with the innovative use of attention mechanisms and style injections, sets a new benchmark in the field of synthetic image generation for time-lapse microscopy.

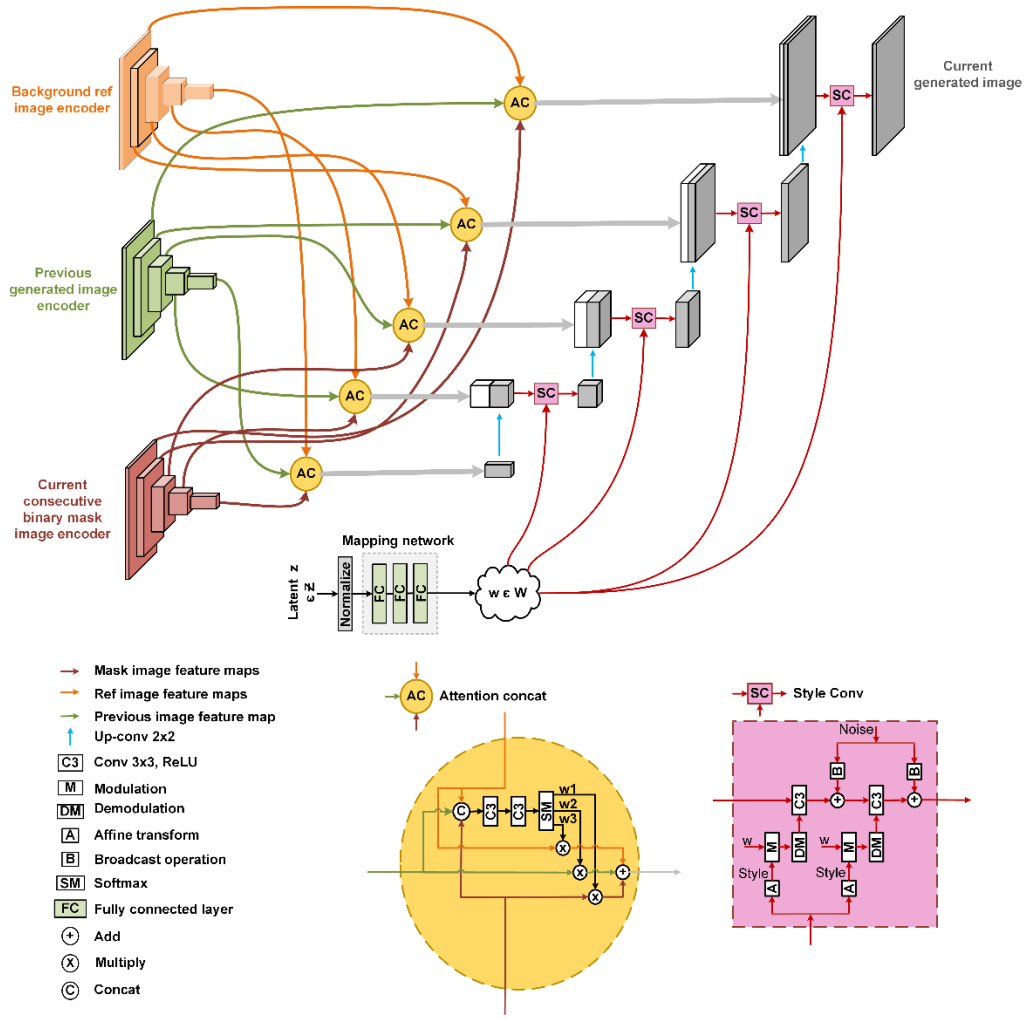

Figure 6.2: Detailed architecture of the low-resolution sequence-based generative model utilizing 2D-UNET encoding and decoding pathways, attention mechanism, and style injections.

6.2.4 Super-resolution image-to-image generative model overview

Following the generation of low-resolution synthetic cell image sequences, our approach employs a super-resolution generative model to further refine and enhance these synthetic images (Figure 6.3). This super-resolution model, similar to its low-resolution counterpart, is built on a UNET-based architecture [207] but with additional enhancements. This model similarly integrates style and noise injection techniques in its decoding path,

which is influenced by StyleNet principles [215]. The style and noise injection significantly enhance the textural details and stylistic elements, contributing to the generation of more realistic and aesthetically consistent images, as demonstrated in our previous research [208]. The architecture features a sequence of encoder blocks that increase feature map depth, capturing intricate cell details. This is followed by a bottleneck process that prepares these features for nuanced reconstruction. In the decoder stages, the model combines upsampled features with style and noise information, progressively enhancing image resolution and quality. An additional upsampling layer in the decoder further increases the output resolution to 512x768, which can be adjusted to higher resolutions by modifying the model parameters.

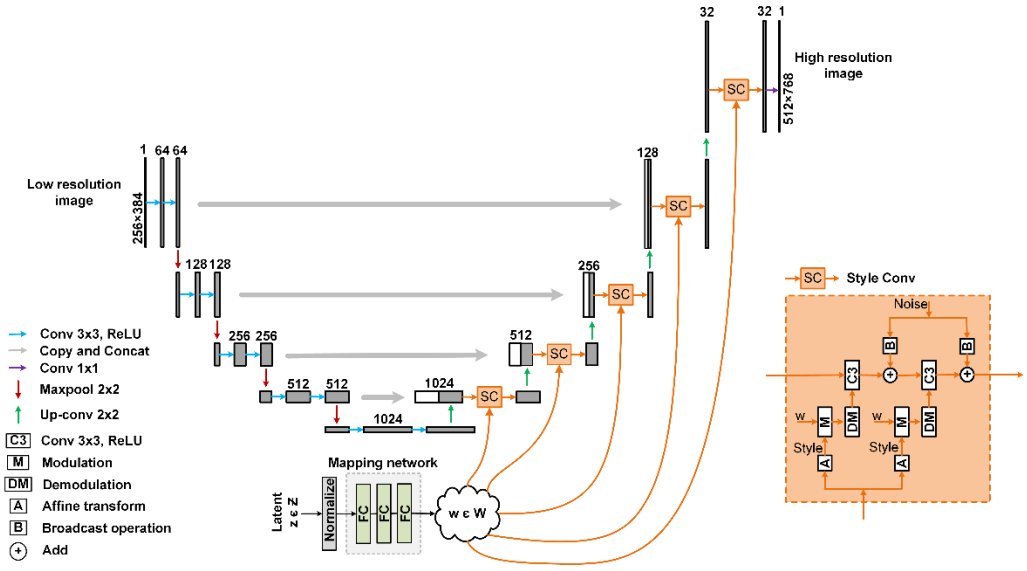

Figure 6.3: Super-resolution generative model architecture for refining low-resolution synthetic cell image sequences, integrating UNET-based design with style and noise injection techniques.

6.2.5 Discriminator architecture overview

The discriminator plays a pivotal role in distinguishing between real and generated images. We have chosen the PatchGAN architecture for the discriminator of both low-resolution and high-resolution training, enhanced with the addition of a linear attention layer in its early layers (Figure 6.4). This architecture and the inclusion of specific components are deliberate choices aimed at optimizing the discriminator's performance. PatchGAN is known for its effectiveness in distinguishing fine details in images, making it a good choice for our purposes [209]. Unlike traditional discriminators that classify an entire image as real or fake, PatchGAN focuses on classifying smaller patches of the image. This approach is particularly beneficial for our model as it ensures that the generated images not only look realistic on a macro scale but also maintain high fidelity in finer details. In the measurement of our low-resolution discriminator loss, we employed a specific approach to enhance the model's discriminative capability. At each training step, we concatenated the last n (e.g., two) real and synthetic (fake) frames along with their corresponding last two real binary mask frames. This concatenation provides the discriminator with a more comprehensive context, allowing it to assess not just the individual frames but also their temporal consistency and alignment with the binary masks. This technique is particularly effective in reinforcing the discriminator's ability to discern subtle differences between real and generated image sequences, thereby sharpening the adversarial dynamic of the model.

The integration of a linear attention layer in the early layers of the discriminator is a beneficial approach. Attention mechanisms have gained popularity in various deep-learning applications for their ability to enhance model performance by focusing on relevant features while ignoring irrelevant ones [210]. The linear attention layer in our model allows the discriminator to prioritize certain aspects of the image, such as specific textures or patterns that are crucial for making accurate classifications. This focused approach

improves the model's efficiency and accuracy in distinguishing real images from synthetic ones.

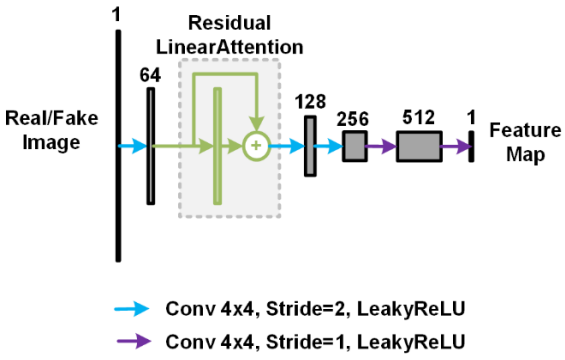

Figure 6.4: Discrimination model architecture employing PatchGAN techniques with linear attention.

6.2.6 FlowNet architecture overview

In the training process of the low-resolution video-to-video generative model, we simultaneously trained a FlowNET model (Figure 6.5), a specialized component designed to calculate and integrate flow loss. This flow loss is crucial for accurately simulating the dynamics of cell movements, thereby enhancing the temporal consistency across video frames [211]. We incorporated this additional loss metric to optimize the generative model's weight updates, specifically aiming to maintain temporal coherence in the synthesized video sequences. This strategy ensures that the generated sequences not only mirror real-world temporal dynamics but also enhance the realism and scientific applicability of the generated images.

The optical flow loss, calculated by our FlowNET, is pivotal in maintaining the integrity of temporal dynamics across generated frames, aligning these dynamics closely with real-world observations to enhance the realism and scientific utility of the synthetic images. It's important to note, however, that the application of flow loss is selectively

applied; it is not utilized in the high-resolution model, which focuses primarily on enhancing image detail through image-to-image translation, thus reducing the need for temporal consistency in that specific context.

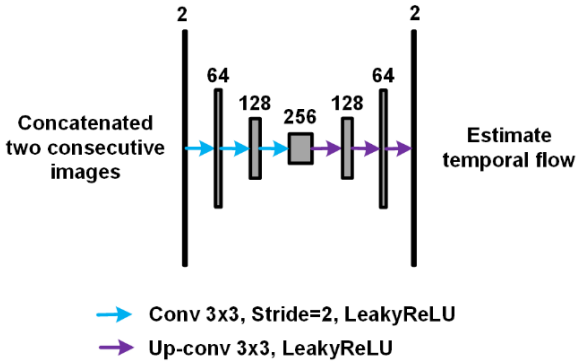

Figure 6.5: Our FlowNET model architecture used in the training process of the video-to-video low-resolution generative model.

6.2.7 Augmentation Process in Training

We employed a robust augmentation pipeline to enhance the training of both the generators and discriminators. This process involves applying a series of video-level augmentation functions to each training sample, designed to introduce variability and improve the model's generalization capabilities. As illustrated in Figure 6.6A, we applied some mostly used conventional image augmentation functions, including random adjustments in histogram equalization, sharpness, brightness, and contrast, as well as horizontal and vertical flips, cropping, saturation modifications, and the addition of Gaussian noise and blur. The training algorithm executes a sequence of the provided augmentation functions for each cell and mask video pair with a pre-defined probability value 'p_vanilla'. In the requested augmentation pipeline, each function is randomly chosen with a consistent probability of 50% and is also applied in a randomized sequence.

In the training of GAN models, the limitations of conventional augmentation become apparent, particularly in its inability to significantly diversify the generated images when the training dataset size is limited. To address this challenge, the concept of differentiable augmentation, detailed in [212], proves to be invaluable. This approach has also been validated in our previous research [208]. Differentiable augmentation applies the same random augmentations to both real and fake samples in a way that is differentiable with respect to the model parameters. This approach encourages the discriminator to mitigate overfitting and improve training stability, making it particularly beneficial for GANs trained with limited data, thus causing the generator to produce more diverse images, thereby improving the overall image generation performance. In the training process, we ran five distinct differentiable video-level augmentation functions, such as random contrast, brightness, cutout, translation, and saturation (Figure 6.6B). The application of each augmentation is controlled by a predetermined probability variable, 'p_diff'. To promote unbiased representation and randomness in the training data, these augmentations are applied in a random sequence. Each function has an equal chance of being selected, set at a 50% probability.

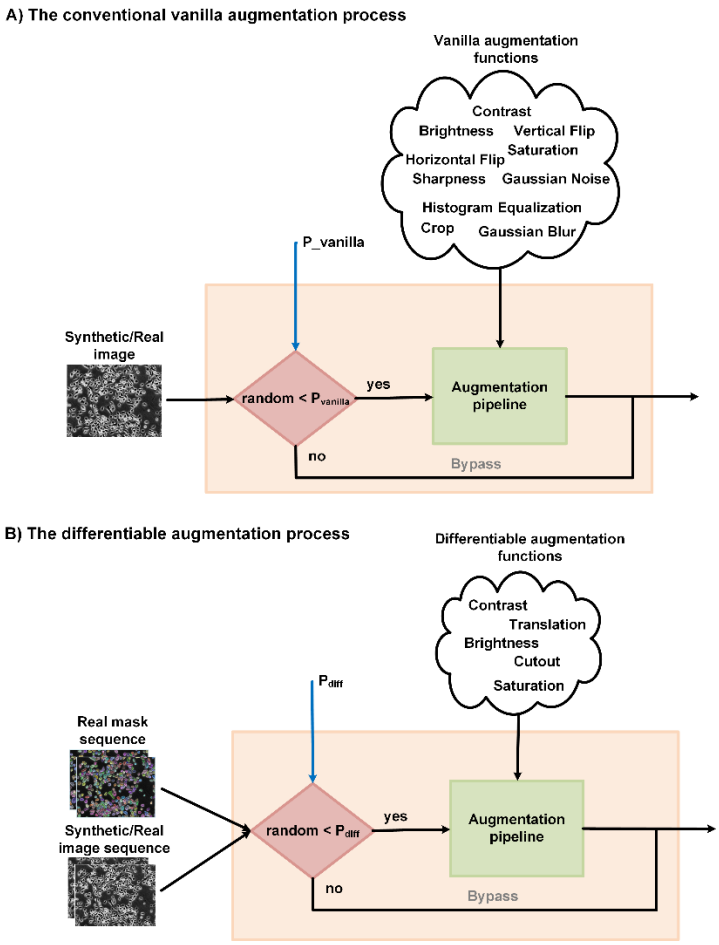

Figure 6.6: The proposed video-level augmentation approach used in the training process.

6.2.8 Loss functions and their rationale

We employed a combination of other critical loss functions (Equations 6.1-6.3) to guide the models effectively during the training process. These include perceptual (VGG) loss [216], L1 loss, and discriminator loss, each serving a specific purpose and contributing to the overall performance and accuracy of the model. The perceptual (VGG) loss and the L1 loss, respectively, ensure perceptual and pixel-wise similarity between the generated and real images. These losses focus on high-level features and granular accuracy. The

discriminator loss, Mean Squared Error (MSE), plays a dual role: for the generator, it gauges effectiveness in deceiving the discriminator, and for the discriminator, it evaluates its capacity to differentiate real from synthetic images. This adversarial loss is key in driving the generative process toward producing images that closely mimic real ones.

We also, in the training process, assigned a specific weight (w1-w3) for each of these loss functions aiming to ensure a balanced contribution during the optimization process. This weighting is crucial as it fine-tunes the impact of each loss function according to its relevance and importance in the image generation task. It is important to note that we do not employ flow loss in the training process of the super-resolution model, as the FlowNET, which calculates flow loss, is not utilized in this phase of the training. This decision is based on the super-resolution model's focus on image-to-image translation rather than temporal video sequence generation.

$$\text{Low\_Res\_Gen\_loss} = \text{D\_MSE} + \text{w1} \times \text{Floss\_Loss} + \text{w2} \times \text{L1\_Loss} + \text{w3} \times \text{VGG\_Loss} \quad (6.1)$$

$$\text{High\_Res\_Gen\_loss} = \text{D\_Fake\_MSE} + \text{w2} \times \text{L1\_Loss} + \text{w3} \times \text{VGG\_Loss} \quad (6.2)$$

$$\text{D\_loss} = 0.5 \times \text{D\_Real\_MSE} + 0.5 \times \text{D\_Fake\_MSE} \quad (6.3)$$

6.2.9 Code availability

The Python scripts that implement the methodologies we developed are publicly available for download at our GitHub repository: https://github.com/abzargar/tGAN. Additionally, the image datasets used in our study can be accessed through a link on the repository page, enabling straightforward replication and further research.

6.3 Results

6.3.1 Model's performance evaluation

In Table 6.1, we compared the performance of our model with the vid2vid model, a notable example among state-of-the-art video-to-video generative models [217]. The vid2vid framework is particularly relevant for our comparative analysis as it is among very few publically available generative models capable of translating mask scene sequences into realistic scene sequences and generating high-resolution video frames. This capability aligns closely with our model's application in generating annotated video datasets, making it an appropriate benchmark for assessing our model's efficacy. we employed a set of metrics to benchmark its performance. This included the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for assessing image quality and similarity to real frames [218]. Additionally, we utilized the Frechet Video Distance (FVD) [219] to measure the distributional similarity between generated and real images, providing insights into the perceptual quality of our model's outputs. To evaluate the temporal coherence of generated video sequences, we adopted specific metrics designed to assess the smoothness and continuity of changes across frames. Finally, the Perceptual Image Patch Similarity (LPIPS) metric [220] was used to gauge the perceptual resemblance of generated images to real ones, ensuring that our model's outputs align closely with human visual judgment. Together, these metrics provided a robust framework for evaluating and benchmarking our model in a video-to-video generation. As presented in Table 6.1, our tGAN model obtained better performance across almost all the metrics and five cell types when tested on the DeepSea [213] and Cell Tracking Challenge [214] dataset time-lapse image sequences. The use of these two annotated time-lapse microscopy datasets, with their challenging imaging conditions, further emphasizes the robustness and versatility of our approach, making it a significant contribution to the field of biomedical imaging and analysis. In Tables 6.2 and 6.3, we further evaluate the quality of video frames generated by our tGAN and

the vid2vid model. This assessment is conducted using the DeepSea cell tracker [213] and Bayesian Tracker (btrack) [223], comparing the results to real annotated time-lapse videos from the DeepSea dataset. The evaluation leverages various object-tracking metrics that we described in our previous research [213], allowing for a detailed comparison of each cell-tracking model's performance. This evaluation focuses on how closely our model's tracking scores align with the ground truth annotated cell image sequences and align with our primary objective, which is to develop a GAN model capable of producing realistic annotated time-lapse microscopy images, addressing the scarcity of annotated data for training the sequence-based deep learning models, such as cell trackers.

The DeepSea tracker, already trained on the DeepSea dataset samples, shows better results than btrack in all cases. More importantly, our tGAN model shows better and closer results to real annotated time-lapse microscopy cell images compared to the vid2vid model. This confirms that the video frame sequences generated by our tGAN model possess more realistic static and dynamic structures across frames, further validating the effectiveness of our approach in producing high-quality synthetic imagery.

Table 6.1: Comparative assessment of model performance. This table presents a comparison of our model's performance across five different cell-type image sequences against the vid2vid model, measuring different video similarity metrics.

Method	FVD (↓)	SSIM (↑)	PSNR (↑)	LPIPS (↓)
Mouse Embryonic Stem Cells [21]				
vid2vid [217]	44.18	0.81	25.31	0.36
tGAN (ours)	8.83	0.95	32.58	0.19
Bronchial Epithelial Cells [213]				
vid2vid	47.32	0.83	26.61	0.35
tGAN (ours)	17.21	0.90	28.42	0.25
Mouse C2C12 Muscle Progenitor Cells [213]				
vid2vid	12.72	0.73	19.19	0.36
tGAN (ours)	14.93	0.89	23.14	0.21
PhC-C2DH-U373 [214]				
vid2vid	8.86	0.88	25.91	0.25
tGAN (ours)	6.43	0.91	26.58	0.18
PhC-C2DL-PSC [214]				
vid2vid	127.58	0.51	10.03	0.39
tGAN (ours)	98.27	0.84	15.68	0.13

Table 6.2: Quality evaluation of synthetically generated time-lapse microscopy sequences using DeepSea Cell tracking model, measuring different object-tracking metrics.

Time-lapse test sequences	MOTA (↑)	MT (↑)	ML (↓)	Precision (↑)	Recall (↑)
Mouse Embryonic Stem Cells [213]					
Real	0.90	0.82	0.02	0.96	0.94
Synthetic Vid2vid	0.73	0.57	0.09	0.94	0.81
Synthetic tGAN	0.93	0.90	0.01	0.97	0.98
Bronchial Epithelial Cells [213]					
Real	0.93	0.94	0.02	0.96	0.98
Synthetic Vid2vid	0.88	0.88	0.03	0.89	0.90
Synthetic tGAN	0.91	0.92	0.03	0.95	0.97
Mouse C2C12 Muscle Progenitor Cells [213]					
Real	0.80	0.64	0.03	0.93	0.89
Synthetic Vid2vid	0.52	0.23	0.25	0.74	0.64
Synthetic tGAN	0.76	0.60	0.06	0.94	0.85

Table 6.3: Quality evaluation of synthetically generated time-lapse microscopy sequences using the btrack model, measuring different object-tracking metrics.

Test set	MOTA (↑)	MT (↑)	ML (↓)	Precision (↑)	Recall (↑)
Mouse Embryonic Stem Cells [213]					
Real	0.85	0.80	0.02	0.93	0.94
Synthetic Vid2vid	0.73	0.51	0.18	0.86	0.84
Synthetic tGAN	0.92	0.90	0.01	0.96	0.97
Bronchial Epithelial Cells [213]					
Real	0.84	0.75	0.25	0.87	0.98
Synthetic Vid2vid	0.78	0.52	0.12	0.79	0.95
Synthetic tGAN	0.83	0.73	0.27	0.86	0.96
Mouse C2C12 Muscle Progenitor Cells [213]					
Real	0.80	0.62	0.04	0.93	0.88
Synthetic Vid2vid	0.53	0.23	0.25	0.74	0.63
Synthetic tGAN	0.75	0.56	0.09	0.93	0.81

Figure 6.7 showcases examples of two consecutive frames generated by our tGAN for each of the three DeepSea cell types. Additionally, these outputs are compared with the corresponding outputs from the vid2vid model. As observed, our tGAN-generated images successfully capture realistic details of the cell bodies, including features like the nucleus, as well as the nuances of the background, demonstrating the model's effectiveness in rendering intricate biological structures. Figure 6.8 also compares the performance of our

tGAN output with the vid2vid model, specifically focusing on the ability to replicate fine visual features of single stem cells. This comparison highlights the intricacies and effectiveness of each model in capturing detailed cellular characteristics. In Figure 6.9, we also present the capability of our proposed approach in generating cell image sequences against a variety of backgrounds given two reference background images. As shown, these backgrounds are precisely referenced from the reference background image used in the low-resolution video-to-video model, highlighting our method's adaptability in replicating diverse cell environmental settings.

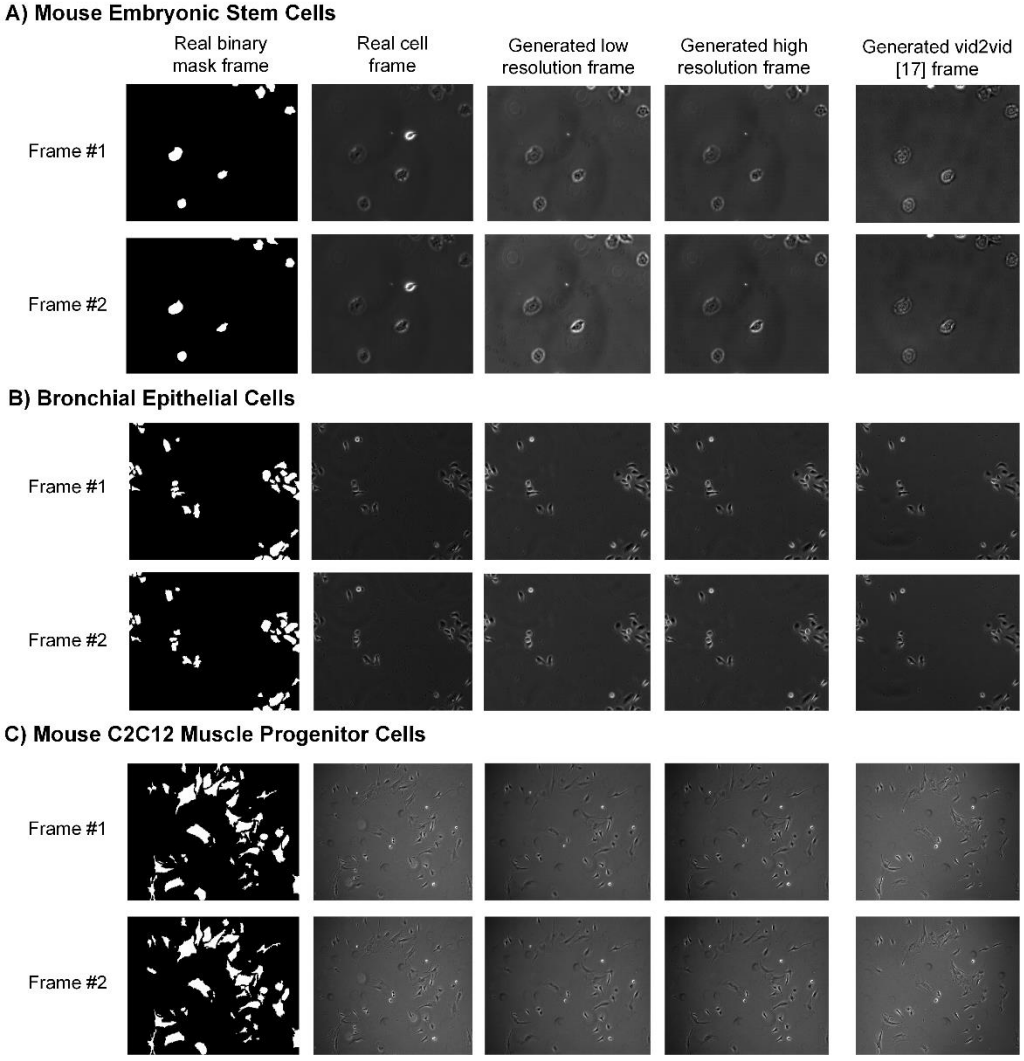

Figure 6.7: The examples of two consecutive synthetic cell images generated by our proposed tGAN model compared to the vid2vido model outputs.

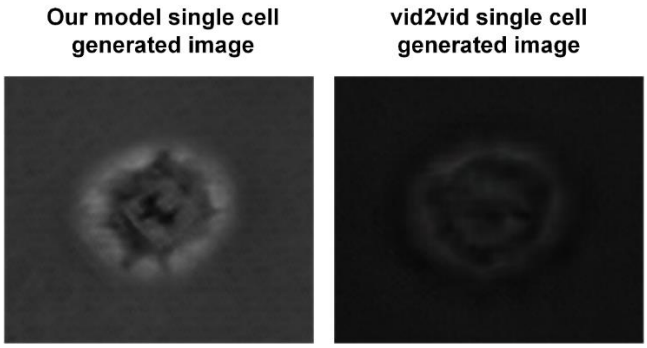

Figure 6.8: Comparative performance of our model vs. vid2vid in replicating fine visual features of single stem cells.

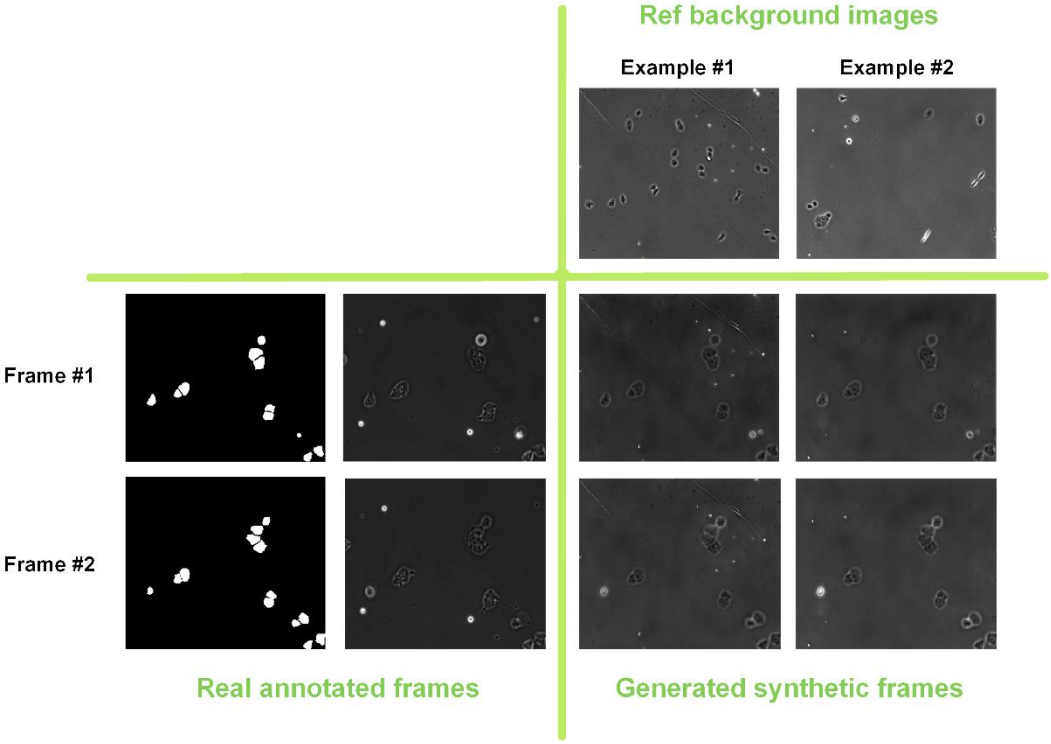

Figure 6.9: The examples of applying different reference background images for the same sample.

In a plot analyzing the FVD score against the length of the tGAN-generated video in frames (Figure 6.10), we observed that the FVD scores exhibited only a slight fluctuation, approximately 1 FVD unit when comparing videos ranging from 10 to 30 frames in length.

This consistency in FVD scores, regardless of video length, underscores our tGAN model's stability and reliability in generating high-quality video sequences over varying durations. Besides, we noticed that the tGAN training dataset (DeepSea and Cell Tracking Challenge Datasets) predominantly contained low and mid-density cell image sequences, which led us to investigate if our model could generate synthetic high-density cell images, a type not seen during training. This aspect is crucial as manual annotation of high-density images is laborious and prone to errors. Therefore, we developed an algorithm specifically for creating synthetic high-density and colony-like time-lapse binary mask images (a task that can be comparatively straightforward using image processing techniques). These were then used as inputs for our tGAN generator during testing. Figure 6.11 demonstrates our model's successful extrapolation from low and mid to high-density images, proving its ability to produce a broad spectrum of realistic cell images and highlighting its potential in various applications.

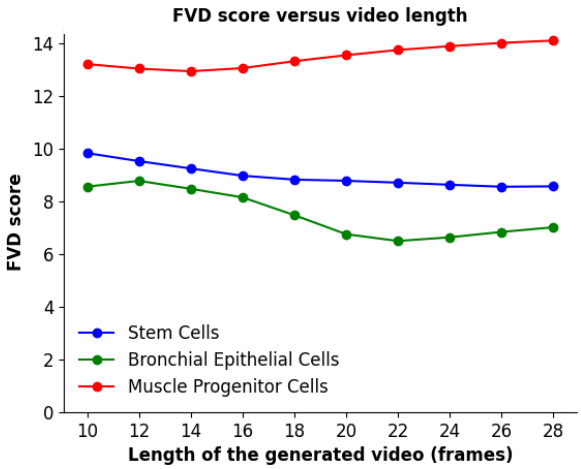

Figure 6.10: Frechet Video Distance (FVD) scores across different video lengths for three DeepSea cell-type time-lapse video frames.

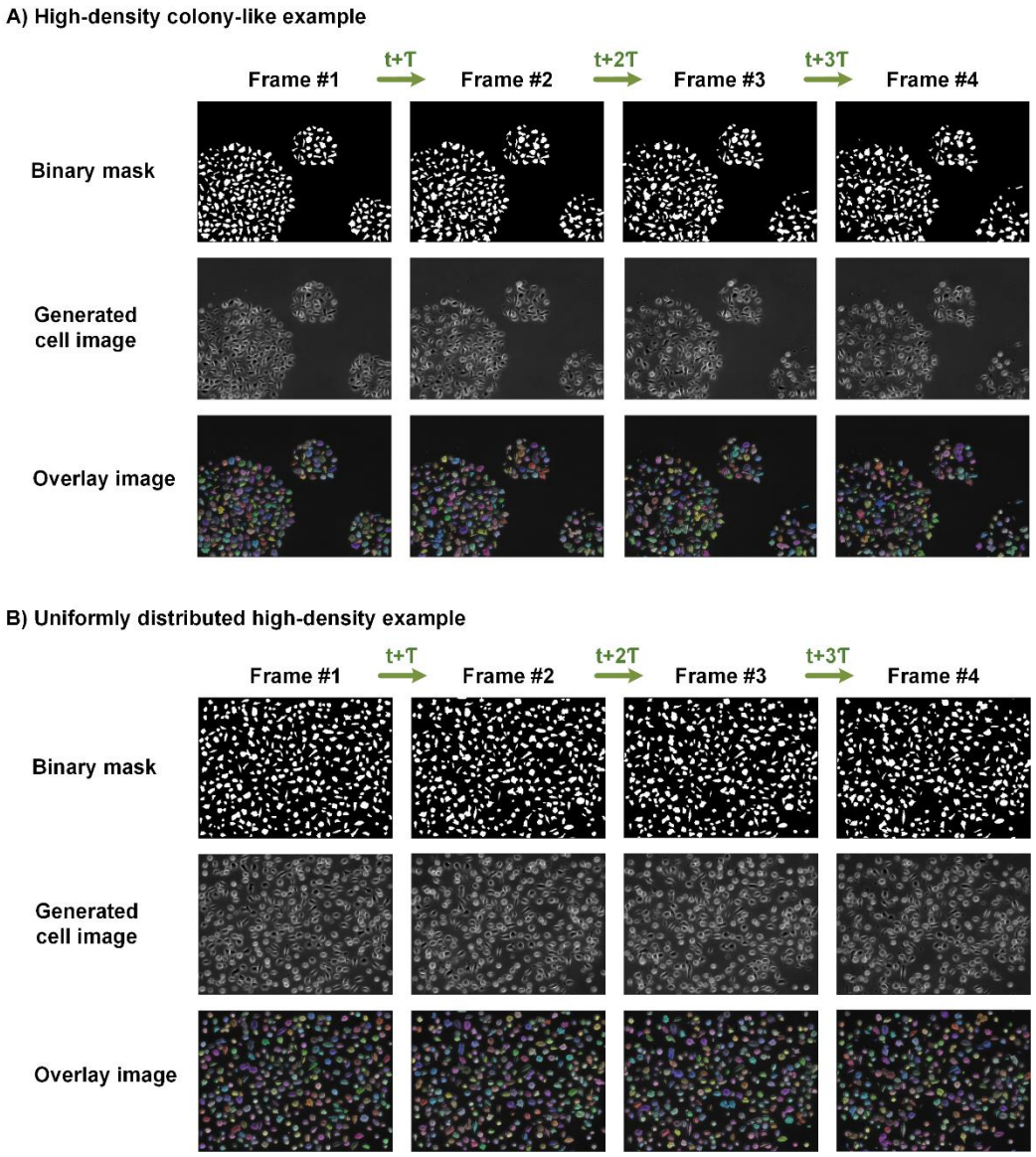

Figure 6.11: The examples of producing synthetic high-density and colony-like time-lapse cell video frames using our tGAN given synthetically generated binary mask images.

6.4 Conclusion

A foundational aspect of developing deep learning models, particularly in the field of biomedical imaging, is the availability of a large and diverse annotated dataset. However, the creation of such datasets for microscopy images is often hindered by the laborious and time-consuming nature of manual annotation. Addressing this bottleneck, our study introduced a GAN-based super-resolution video generator designed to bridge the gap caused by the scarcity of annotated microscopy datasets, a field that demands high accuracy and detail in image processing. By utilizing our model, we effectively circumvent the need for extensive manual annotation, generating realistic and diverse sets of synthetic annotated time-lapse cell images. This approach not only enhances the breadth of available training data but also significantly boosts the performance of the models, such as the cell segmentation and cell dynamic tracking models, particularly in scenarios with limited annotated datasets.

The model's two-part structure adeptly handles both low-resolution and high-resolution image generation, a design choice that has proven pivotal in its performance. The integration of style and noise injection, coupled with the inclusion of a reference background image, improves the generation of synthetic time-lapse cell images. These features not only enhance the realism and detail of the generated images but also ensure their adaptability to diverse environmental settings. A critical aspect of our study was the use of different cell types, encompassing a variety of challenging cell imaging scenarios. Our model's superior performance in generating high-quality, realistic time-lapse microscopy videos, as evidenced by its outperformance of state-of-the-art models, underscores its effectiveness in handling complex microscopy images. This success is quantitatively supported by comprehensive metrics, which collectively affirm the model's superiority in image quality, temporal coherence, and perceptual accuracy. Besides, our

model's ability to generate high-density cell images marks a useful advancement in the field. This capability addresses a critical need in biological research, where high-density cell images are often challenging to annotate manually. By generating synthetic high-density cell binary masks and using them as the model inputs in the test phase, we demonstrated the model's capacity to extend its application beyond the conditions experienced during training. This extrapolation is a practical benefit that can help reduce the time and effort required for manual annotation.

In conclusion, our tGAN model stands as a robust tool in biomedical imaging, capable of generating high-quality synthetic annotated time-lapse cell images across a spectrum of densities and cell types. While our current focus has been on the DeepSea and Cell Tracking Challenge datasets, the model's architecture and performance suggest a broader applicability across various cellular imaging modalities. Future work could explore the model's adaptability to even more diverse cell types and imaging conditions, as well as optimize computational efficiency for larger datasets. The potential for further enhancing the model's generalization capabilities remains an exciting prospect, paving the way for its application in a wider array of biomedical research scenarios.

7 Quantitative Analysis of Single-Cell Images

7.1 Introduction

In recent years, the field of cellular biology has been transformed by advances in imaging and computational technologies, enabling researchers to study single cells in unprecedented detail [222, 223]. This chapter focuses on the quantitative analysis of single cells, employing cutting-edge imaging techniques and sophisticated data analysis methods to delve into the cellular microcosm. Our primary goal is to characterize the inherent heterogeneity and dynamic behaviors of cells, which are often obscured in bulk cell analyses.

Single-cell analysis is pivotal for understanding the complex landscape of cellular functions across different cell types and states. By quantitatively measuring a wide range of cellular features such as shape, size, and internal structures, we can glean insights into the roles individual cells play in both health and disease. This approach is particularly beneficial for identifying subtle phenotypic variations that contribute to diverse biological functions and responses to environmental stimuli [224, 225].

Moreover, the ability to dissect the interactions of cells within their microenvironment opens new avenues for exploring how cells communicate and cooperate with each other, and how they contribute to the architecture and functionality of tissues. These insights are crucial for unraveling the mechanistic bases of biological processes and for understanding the progression of complex diseases such as cancer, where the

microenvironment plays a key role in disease development and response to treatment [226].

In addition to providing a snapshot of cellular diversity, quantitative single-cell analysis facilitates the study of cellular dynamics over time. By tracking changes in individual cells, we can observe cellular responses to external stimuli, understand cellular aging processes, and investigate mechanisms of disease progression at a granular level. This dynamic perspective is essential for developing more effective therapeutic strategies that target specific cell populations or that modulate the cellular environment [227, 228].

In this chapter, we describe the use of computational methods to conduct a quantitative analysis of single cells from the DeepSea dataset [229]. We discuss the computational tools and algorithms used to extract meaningful data from single-cell images, highlighting how different cell types exhibit unique morphological and functional traits. Initially, high-resolution microscopy images of DeepSea stem cells, lung cells, and muscle cells were segmented using the DeepSea tool [229], ensuring accurate cellular representations. Following segmentation, detailed feature extraction was performed using both conventional morphological metrics and innovative pixel intensity-based features. Morphological data was further processed with the Celltool software for alignment and normalization, facilitating comparative analyses across different cell types and conditions.

In addition to morphological analysis, we extracted spatial and frequency features to explore more granular aspects of cell structure and function. These features were analyzed to understand the distribution and variability in pixel intensity, which reflect underlying cellular processes not apparent through morphology alone. We applied statistical methods and machine learning algorithms, including k-means clustering and silhouette score analysis, to assess the clustering quality and distinguishability of the features. This comprehensive approach enabled us to characterize cellular heterogeneity

and uncover critical insights into the dynamic behaviors and interactions of single cells within their microenvironments.

7.2 Methods

7.2.1 Dataset

In this study, we leverage our DeepSea dataset [229], which consists of high-resolution time-lapse microscopy images capturing three distinct cell types: mouse embryonic stem cells, bronchial epithelial cells, and mouse C2C12 muscle progenitor cells. The inclusion of time-lapse imaging enables the study of temporal changes within cells, offering insights into cellular processes over time. Utilizing this dataset, our research aims to dissect the complex biological characteristics of these cell types, understanding their unique and shared features across different conditions and over time. This comprehensive approach ensures a deep and nuanced exploration of cellular morphology and behavior, critical for advancing our understanding of cell biology.

7.2.2 Cell Shape Alignment and Normalization Using Celltool

The morphological analysis of single cells commenced with an in-depth examination of cell shapes. To ensure the integrity of our results, it was necessary to perform precise cell shape alignment and normalization. The individuality of each cell's orientation, specifically the variance in the major axis across the population, presents a significant analytical challenge. To facilitate comparative analysis and feature extraction, standardizing these orientations was imperative [230, 231].

Celltool [232], a comprehensive suite tailored for cellular morphology analysis, was instrumental in our process. This robust collection of analytical tools allowed us to meticulously process the segmented cell images. Its functionality extends beyond mere extraction; Celltool systematically aligns every cell contour along its longest axis (Figure 7.1). This standardization is critical for establishing a uniform frame of reference across all

cells, thereby enabling the extraction and comparison of shape features on a common scale.

With the assistance of Celltool, we could navigate the complexities of cellular morphology. By plotting the variation in cell shape—referred to as shape modes—we were able to visualize and quantify the degree of morphological diversity present within our cell populations. The software's capability to perform statistical comparisons of shape distributions empowered us to discern subtle morphological differences and categorize cells into distinct phenotypic profiles.

This systematic approach to cell shape alignment and normalization not only standardized the orientation and scale of the cells but also provided a foundation for a more nuanced and quantitatively robust analysis. As a result, the extracted shape features could be directly correlated to cellular function and phenotype, offering a clear window into the biological significance of morphological variation. This preparatory step, therefore, was pivotal in achieving the overarching goal of our study: to elucidate the complexities of cellular form and function through quantitative analysis.

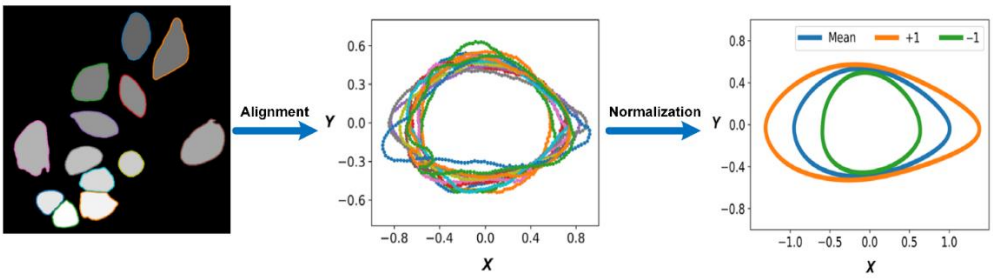

Figure 7.1: Cell shape alignment and normalization using Celltool [232].

7.2.3 Cell Feature Extraction

- **Cell Shape Features**:

The first category we delved into was cell shape features. These attributes are crucial as they often correlate directly with the specialized functions cells perform within the body. Evolution has sculpted cells into specific shapes to optimize their roles, from nutrient absorption to rapid mobility. Therefore, by examining the shape of cells, we gain valuable insights into their biological roles.

To obtain these shape features, the initial step is to distinguish the cell body from its surroundings using segmentation tools like DeepSea. Once isolated, a variety of image processing techniques can be applied to extract a suite of shape features (Figure 7.2). In our study, we identified and computed 20 significant shape features that include area, perimeter, aspect ratio, rectified area, extent, convex area, solidity, equivalent diameter, orientation, major and minor axis lengths, eccentricity, axis ratio, circularity, convex perimeter, waviness, and the Fréchet extrema and coefficient of variation (fret_max, fret_min, fret_mean, fret_CV). Each of these features offers a different perspective on the cell's morphology and, by extension, its function.

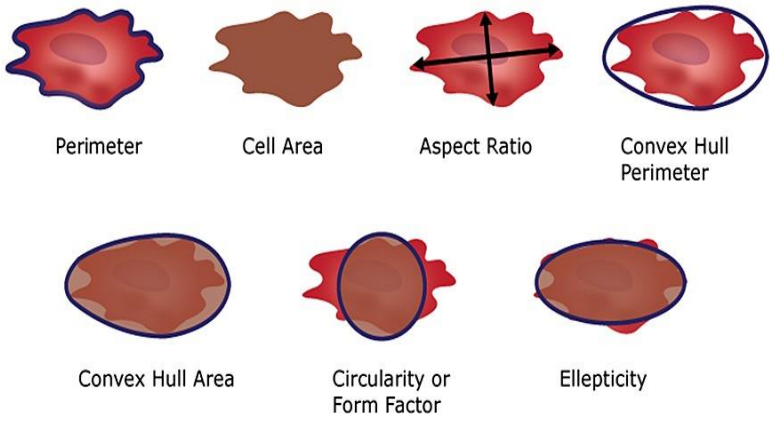

Figure 7.2: Some of the cell shape features we extracted from the segmented cell body.

- **Frequency and Spatial Features**:

While shape features are visually intuitive and easily interpretable, they represent only a portion of the information available within cellular images. To broaden our analysis, we incorporated frequency and spatial features, which can be thought of as 'hidden' features. These attributes are not immediately apparent to the naked eye but can be critical indicators of cellular behavior and function.

As illustrated in Figure 7.3, the spatial features we analyzed, such as texture, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Difference Method (GLDM), give us an understanding of the distribution of pixel intensities across the cell's area [233, 234]. In contrast, frequency features derived from the Fast Fourier Transform (FFT) and wavelet transform examine the variation in pixel intensity across the cell's domain [235, 236]. These features provide a more comprehensive view of a cell's internal structure and activity by detailing how often pixel intensities fluctuate in various directions.

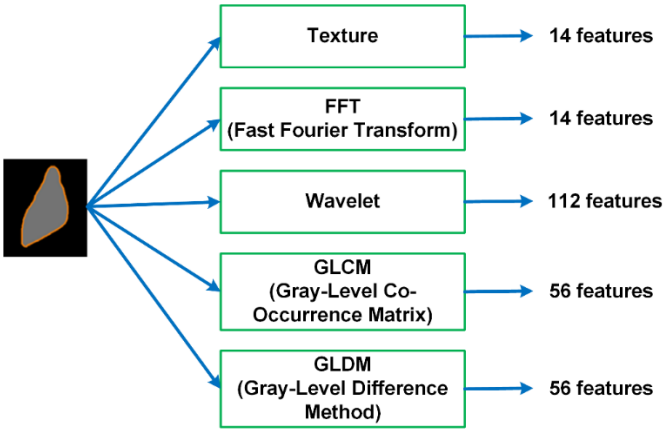

Figure 7.3: Frequency and spatial features we extracted from the segmented cell body.

- **Comprehensive Feature Analysis:**

Upon extraction, we were faced with a rich dataset comprising 251 frequency and spatial features alongside the 20 cell-shape features. The next step was to evaluate these features to determine their significance and ability to convey cellular information. Our goal was to identify the most salient and distinctive features—those that not only define cellular identity but also carry implications for understanding cell functionality.

This endeavor provided us with a multi-faceted view of cells, combining the tangible aspects of cell shape with the more elusive but equally important frequency and spatial features. By doing so, we aimed to capture a holistic picture of the cell, one that extends beyond mere form to encompass dynamic processes and interactions within the cell's internal and external environments.

7.2.4 Cell Feature Clustering and Analysis

Upon extracting a comprehensive set of cellular features, our analysis progressed to the evaluation of their discriminatory power using clustering techniques. Clustering scores serve as a quantitative benchmark for determining the effectiveness of each feature in differentiating individual cells. To facilitate this, we employed the k-means clustering algorithm [237], a widely used method due to its efficiency and simplicity in grouping data into coherent subsets.

The k-means algorithm was utilized to categorize the cell bodies into distinct clusters based on the plethora of extracted features (Figure 7.4). This unsupervised learning algorithm operates by partitioning the cells into k clusters, where each cell belongs to the cluster with the nearest mean feature vector, thus minimizing within-cluster variances. The process iteratively assigns cells to clusters based on their Euclidean distances from the centroids of the clusters.

To evaluate the performance of the k-means clustering, we relied on the silhouette score—a metric that assesses the quality of the clustering. The silhouette score is calculated for each sample and provides a measure of how similar that sample is to samples in its own cluster compared to samples in other clusters. Scores close to 1 suggest that samples are far away from neighboring clusters, indicating well-separated and defined groupings. Conversely, scores near -0 indicate overlapping clusters with poor separation.

The silhouette score is particularly valuable as it incorporates both cohesion and separation—two key aspects of successful clustering. Cohesion measures how closely related objects are within the same cluster, while separation quantifies the distinctness of different clusters from one another. A high average silhouette score across all cells suggests that the clustering method has successfully captured the intrinsic patterns and structures present in the cellular feature space.

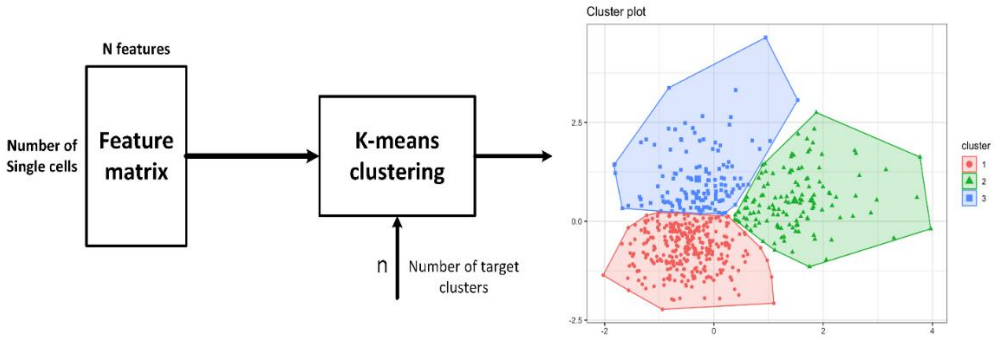

Figure 7.4: Single-cell feature clustering using the K-means method.

7.2.5 Optimal Feature Selection

In the pursuit of refining our feature set for maximal interpretability and efficiency, we adopted Particle Swarm Optimization (PSO) [238] as a heuristic approach to identify

the most impactful and relevant features that contribute to the highest clustering (silhouette) score. PSO is a computational method inspired by the social behavior of birds flocking or fish schooling and is adept at solving complex optimization problems by iteratively improving a candidate solution with regard to a given measure of quality.

Our method integrated PSO to adjust importance weights dynamically within the original cell feature vector (Equations 7.1 and 7.2). As depicted in the accompanying flowchart in Figure 7.5, the PSO algorithm navigates through the multi-dimensional feature space, assigning and adjusting weights with the explicit goal of maximizing the silhouette score. Throughout this optimization process, the algorithm evaluates the contribution of each feature to the clustering outcome.

$$F_{Original} = [F_1, F_2, F_3, \ldots, F_n] \tag{7.1}$$

$$F_{New} = [w_1F_1, w_2F_2, w_3F_3, \ldots, w_nF_n] \tag{7.2}$$

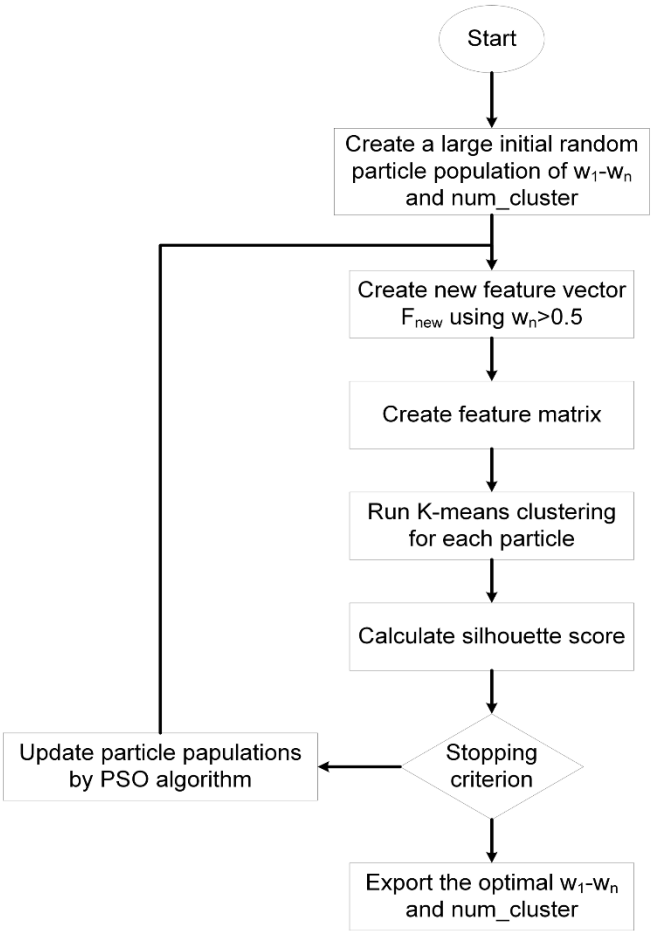

Figure 7.5: Single cell feature selection using particle swarm optimization.

One of the defining steps in our PSO application is the iterative selection and deselection of features based on their weighted importance. Features that receive a weight above a pre-defined threshold (in our case, 0.5) are retained, whereas those with lesser weights are deemed less relevant for clustering and are thus excluded from the feature vector. This process results in a dual-fold optimization wherein features are not only selected but also weighted according to their contribution to the overall clustering efficacy.

The PSO-driven selection method culminated in a distilled set of features that hold the greatest significance for distinguishing between different cell populations. This targeted feature set not only enhances the clarity of our clustering analysis but also paves the way for more nuanced biological interpretations. By focusing on features with high-importance weights, we ensure that the selected attributes have a substantial influence on the silhouette score, thereby aligning our analysis closely with the underlying biological phenomena that dictate cellular diversity and behavior. This approach harmonizes feature selection with feature weighting, creating a robust framework for identifying the salient features that capture the essence of our cellular data.

7.3 Results

7.3.1 Cell cycle duration is adjusted based on birth size: showcasing the application of the DeepSea

In this section, we showcased the application of DeepSea by investigating cell size regulation in mESCs across hundreds of cell division cycles. Cells need to grow in size before they can undergo division. Different cell types maintain a fairly uniform size distribution by actively controlling their size in the G1 phase of the cell cycle [239]. However, the typical G1 control mechanisms of somatic cells are altered in mESCs [240, 241]. Mouse embryonic stem cells have an unusually rapid cell division cycle that takes about 10 hours to complete (Figure 7.6A). The rapid cell cycle of mESCs is primarily due to an ultrafast G1 phase that is about 2h compared to ~10 in skin fibroblast cells with daughter cells born at different sizes (Figure 7.7A, B). An interesting question is whether mESCs can employ size control in their rapid G1 phase, just as most somatic cells do.

Using confocal microscopy, we showed that the area of a cell is closely correlated with the cell volume, making the area a faithful measurement of cell size (Figure 7.7C, D). By measuring the size of the sister cells at birth, we showed that 42% of divisions resulted

in daughter cells of different sizes (Figure 7.6B). We hypothesized that smaller-born cells would spend more time growing compared to their larger sister cells. In support of this hypothesis, we observed that smaller-born cells increase their cell cycle duration by about ~2h compared to their larger sister cells (Figure 7.7C).

Together, our results show that DeepSea can be applied to accurately quantify the cell biological features of cells, such as cell size or cell cycle duration. In addition, our findings support the hypothesis that mESCs can adjust the cell cycle duration based on birth size, suggesting cell size control through an unknown molecular mechanism [239]. Besides, it shows that DeepSea can capture cell size distribution that is closer to the ground truth as determined by manually segmented cells.

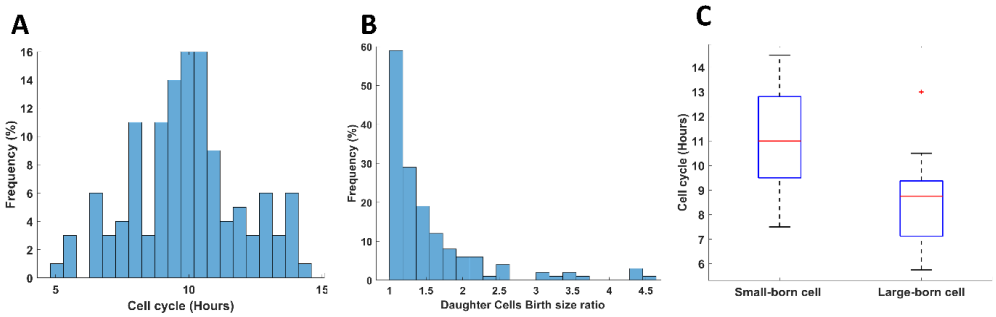

Figure 7.6: Showcasing the DeepSea application. Cell size regulation in mouse embryonic stem cells. (A) Distribution of the cell cycles. (B) Histogram of birth size ratio of daughter cell pairs. (C) Comparing the cell cycle duration of the cells born small with those born large.

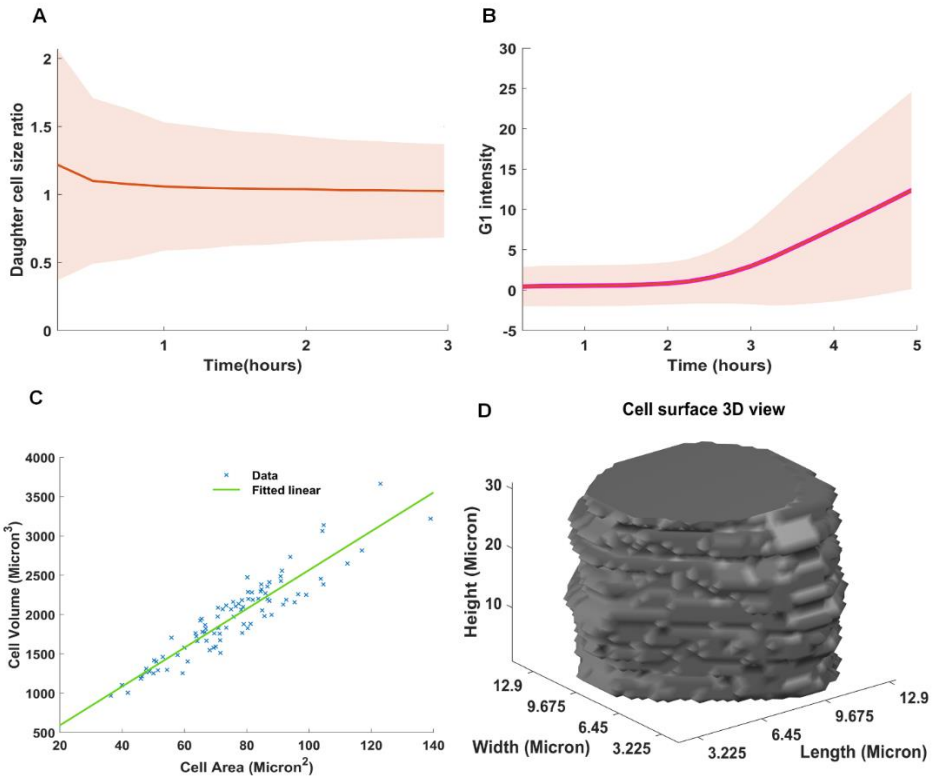

Figure 7.7: Stem cell feature analysis. (A) Cell size ratio graph of daughter cell pairs. (B) Automated measurement of G1 duration using Fucci sensor (Geminin-GFP) that increases its activity as cells enter the S phase. (C) Cell area versus cell volume measurement using confocal microscopy for each embryonic stem cell. (D) One example of cell surface measurement, obtained from our confocal microscopy experiment.

7.3.2 Morphological Diversity Across Cell Types

The application of Celltool to our dataset unveiled a tapestry of shape modes that reflect the intricate morphology of various cell types. As shown in Figure 7.8, bronchial epithelial cells, with their six identified shape modes, predominantly exhibited one primary model (83%) and a secondary model (7%), indicating a dominant structural phenotype. Stem cells presented a simpler morphological spectrum with two shape modes aligning with their more uniform functions. Here, 90% of stem cells conformed to the first shape mode, emphasizing their consistent morphology, while the rest displayed the second mode,

indicating minimal variation. Muscle cells, conversely, demonstrated a broader array of six-shape modes, suggesting a higher degree of morphological complexity and adaptability.

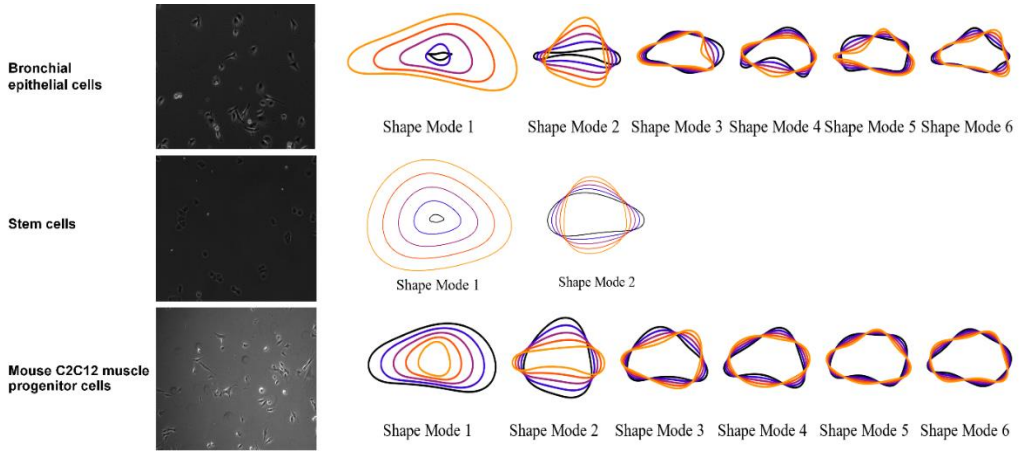

Figure 7.8: Shape mode distribution of three DeepSea dataset cell types obtained using Celltool.

7.3.3 Genotypic Influence on HBEC3kT Cell Morphology

We made a discovery when examining a separate, genotype-specific dataset of HBEC3kT, an immortalized human bronchial alveolar cell line. Using DeepSea software, we segmented this new set of cell lines (3000 cell images), and utilizing Celltool, we identified five distinct shape models and observed notable differences in shape distribution across six genotypic categories that were annotated already. As show in Figure 7.9, The contrast was particularly striking between the ‘empty’ genotype (A3 and B3) and the other four (A1- WT1_Kv12, A2- WT1_LACZ, B1- MUT1_Kv12, B2-MUT1_LACZ), with the latter groups primarily exhibiting the first shape model, suggesting a genotype-influenced structural pattern.

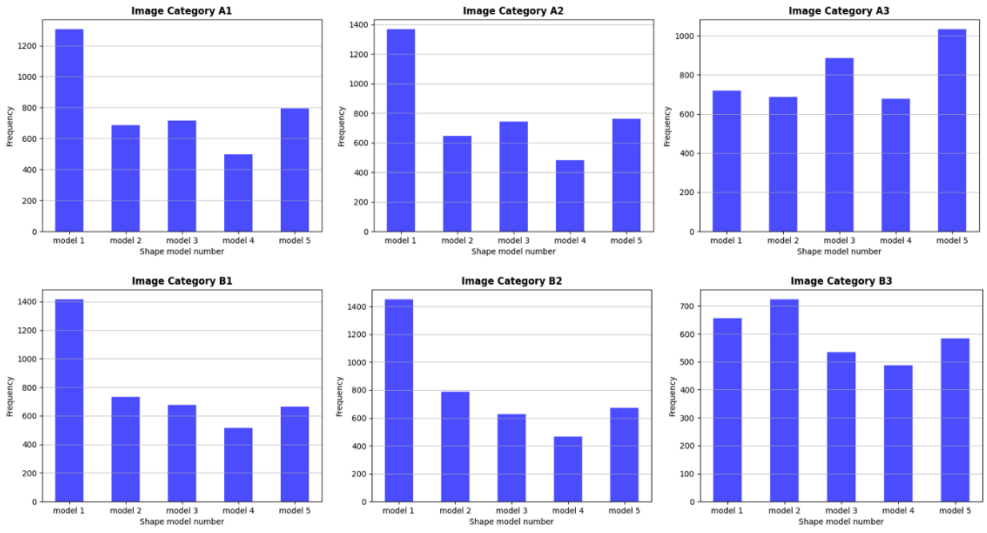

Figure 7.9: Influence of specific genotypes on cell shape patterns.

7.3.4 Feature Clustering Scores and Significance

Employing k-means clustering (with three clusters), we analyzed individual feature clustering scores and prioritized them accordingly. The emergent pattern underscored the predominance of spatial features over shape attributes, with the notable inclusion of shape circularity for stem cells among the top 20 features. This finding reiterates the concept that critical cellular features, often concealed from direct observation, can be pivotal in characterizing cell types.

7.3.5 Comparative Clustering Efficacy

Our clustering score comparisons revealed that spatial features, particularly GLCM and GLDM, consistently achieved higher scores across all cell types, underscoring their discriminating power. Moreover, an experiment aggregating the top 20 stem cell features for clustering resulted in an impressive silhouette score. Interestingly, selecting features with a silhouette score above 60% yielded a substantial score, indicating that a subset of

high-scoring features could provide greater discriminatory power than the entire feature set or solely shape-based features.

7.3.6 Optimization and Clustering with PSO and K-Means

In a further experiment, we applied PSO in conjunction with l-means clustering (choosing two clusters) to refine the selection of shape features. This resulted in a respectable silhouette score of 0.52, with PSO selecting 8 out of the 20 shape features. The top-weighted features—minor axis, aspect ratio, and Feret mean—played a crucial role in defining the clusters. Illustrative cell shape examples demonstrated clear differences between the clusters in these features, justifying their segregation into distinct groups.

Expanding to three clusters slightly reduced the silhouette score to 0.47, with PSO selecting 12 features. The heaviest weights were given to circularity, orientation, and equivalent diameter, reflecting their significance in the clustering process. The provided cell shape examples for these clusters showcased their distinct circularities and orientations, validating the PSO's feature selection in capturing the essence of cell morphology.

These results encapsulate the efficacy of computational tools in deciphering cellular complexity. They also highlight the potential of integrating feature selection algorithms like PSO to enhance the precision of morphological analysis, providing deeper insights into the cell's structural attributes as they relate to function and genotype.

7.4 Conclusion

We showcased the application of DeepSea by investigating cell size regulation in mESCs across hundreds of cell division cycles. Our cell size analysis revealed that smaller-born mESCs regulate their size by spending more time growing in the G1 phase of the cell cycle. These findings strongly support the idea that mESCs actively monitor their size,

consistent with the presence of size control mechanisms in the short G1 phase of the embryonic cell cycle.

This chapter has delved into the intricate world of single-cell analysis, an endeavor made increasingly vital and feasible by significant advancements in imaging and computational technologies. Our multifaceted approach to characterizing the complex heterogeneity and dynamics of single cells has offered an unprecedented window into the cellular microcosm. Through meticulous data collection, segmentation, and the application of Celltool for shape alignment and normalization, we have established a framework for quantitative cellular analysis that underpins the vast potential of this field.

The methods we employed served to distill a vast array of cellular features into meaningful data, enabling us to transcend the traditional boundaries of visual interpretation. Our analyses highlighted the significant roles of both morphological and hidden frequency and spatial features in capturing the full spectrum of cellular diversity. The innovative use of k-means clustering and silhouette scores provided quantitative validation of our feature extraction, emphasizing the remarkable discriminative power of spatial features in particular.

Our exploratory journey into feature selection was further enhanced by the adoption of Particle Swarm Optimization (PSO), which not only refined the selection process but also provided weighted importance to the most critical features. This selective weighting illuminated the path to understanding the functional significance of each cellular attribute, aligning our computational analysis closely with biological relevance.

The results we achieved are emblematic of the synergy between computational prowess and biological inquiry. We observed distinct morphological profiles across lung, stem, and muscle cells, which were intricately linked to their functional roles and genetic

backgrounds. The application of PSO with l-means clustering revealed a nuanced picture of cell morphology and provided a quantitative foundation for categorizing cells based on structural traits.

In conclusion, our work underscores the transformative impact of quantitative single-cell analysis in biological research. By marrying advanced computational techniques with cellular biology, we have carved out a detailed landscape of cellular behavior, functionality, and interaction. The insights gleaned from our study not only contribute to the scientific understanding of cellular diversity but also hold the promise for the development of novel diagnostic tools and targeted therapies. The methodologies and findings presented herein lay a solid groundwork for future explorations, poised to unravel the complexities of life at the single-cell level.

8 Conclusion

This dissertation has journeyed through the intricate dynamics of cellular processes, highlighting the pivotal roles of cell cycle dynamics, live-cell imaging, and the transformative impact of artificial intelligence and machine learning in microscopy. This conclusion synthesizes the key findings and lays out prospective directions for future research that could continue to advance our understanding and application of these fundamental insights.

8.1 Summary

The exploration began with Live-cell imaging in Chapter 2, introducing it as a revolutionary technique allowing real-time observation of cellular processes. This chapter highlighted advancements from traditional microscopy to high-resolution methods like fluorescence and confocal microscopy, which have enhanced our ability to observe cellular functions in their native states without disruption. The significance of these techniques extends beyond basic research, offering profound benefits in areas such as drug development and the personalized medicine approach to treatment, where understanding individual cellular responses is key.

In Chapter 3, the focus shifted to the integration of artificial intelligence and machine learning with microscopy image analysis, demonstrating how these technologies have begun to redefine the possibilities within biological research. AI and ML not only improve the efficiency and accuracy of image analysis but also open new avenues for exploring cellular mechanisms at unprecedented scales and complexities. This integration presents exciting opportunities in diagnostic and therapeutic contexts, particularly through enhanced image resolution and the automated interpretation of complex biological data.

Chapter 4 introduced DeepSea, a versatile and trainable deep-learning model for single-cell segmentation and tracking in time-lapse microscopy images. This model automates the analysis of cellular features and provides insights into cellular dynamics and heterogeneity, achieving high accuracy and robustness in segmentation and tracking tasks.

Chapter 5 presented cGAN-Seg, a CycleGAN-based segmentation model designed to enhance cell segmentation performance with limited annotated training data. This approach overcomes the limitations of existing segmentation models by generating realistic and diverse synthetic microscopy images, significantly improving the segmentation accuracy.

Chapter 6 described tGAN, a GAN-based super-resolution video-to-video generative model for time-lapse microscopy. This model generates high-quality synthetic annotated datasets, facilitating the training of deep learning models for various cell analysis tasks, and enhancing the breadth of available training data.

Chapter 7 showcased the application of the proposed AI tools in quantitative single-cell analysis, investigating cell size regulation in mouse embryonic stem cells (mESCs) and highlighting the significant roles of both morphological and hidden frequency and spatial features in capturing cellular diversity. The comprehensive framework for quantitative cellular analysis provided valuable insights into cellular heterogeneity and dynamics.

8.2 Future Work

The findings and methodologies presented in this dissertation open several promising avenues for future research in live-cell imaging and the integration of AI and ML in microscopy. To build on this work, the following directions are proposed:

1- Expansion of Deep Learning Models for Diverse Imaging Modalities: While this dissertation focused on phase-contrast microscopy and specific cell types, future research could extend the DeepSea, cGAN-Seg, and tGAN models to other imaging modalities such as fluorescence microscopy, confocal microscopy, and 3D imaging techniques. This expansion would enhance the applicability and robustness of these models across a broader range of biological contexts and imaging conditions.

2- Development of Multimodal Data Integration Techniques: Integrating data from different imaging modalities and combining them with other types of biological data, such as genomics, proteomics, and metabolomics, could provide a more comprehensive understanding of cellular processes. Future research could develop AI models that can seamlessly integrate and analyze multimodal data, leading to new insights into cellular mechanisms and disease pathways.

3- Real-Time and Adaptive Imaging Analysis: Enhancing live-cell imaging capabilities to include real-time and adaptive analysis using AI and ML could significantly improve our understanding of dynamic cellular behaviors. Future work could focus on developing models that not only analyze images in real-time but also adapt imaging parameters based on ongoing analysis, optimizing the quality and relevance of the data being collected.

4- Advancements in Cell Tracking and Lineage Tracing: The DeepSea model has demonstrated promising results in cell tracking, but further improvements are needed to track cells through complex behaviors such as cell division, differentiation, and migration in heterogeneous populations. Future research could explore advanced tracking algorithms and lineage tracing methods that provide more detailed insights into cellular development and interactions over time.

5- Enhanced Synthetic Data Generation and Validation: The use of GAN-based models like cGAN-Seg and tGAN for synthetic data generation has shown great potential. However, further refinement is needed to ensure that synthetic data closely mimics real biological conditions. Future work could focus on developing more sophisticated validation techniques to ensure the quality and reliability of synthetic data, as well as exploring the use of synthetic data in training models for rare or difficult-to-obtain cell types.
6- Automated High-Throughput Screening: The integration of AI and ML with live-cell imaging can be extended to automated high-throughput screening of drug effects, genetic perturbations, or environmental changes in cellular behavior. Developing systems that combine automated imaging, analysis, and experimental manipulation would significantly accelerate the discovery of new therapeutic targets and improve our understanding of cellular responses.
7- Collaborative and Open-Source Platforms: To maximize the impact of AI and ML in microscopy, developing collaborative and open-source platforms that allow researchers to share models, datasets, and analysis tools is essential. Future work could focus on creating and maintaining such platforms, fostering a collaborative environment that accelerates innovation and reproducibility in the field.

In conclusion, the advancements presented in this dissertation lay a foundation for future research in live-cell imaging and the integration of AI and ML in microscopy. By pursuing these proposed directions, researchers can further enhance our understanding of cellular processes and develop new tools and methodologies that will drive significant progress in biomedical research and clinical applications.

Bibliography

[1] K. Ji, M. Sameni, K. Osuala, K. Moin, R. R. Mattingly, and B. F. Sloane, "Spatio-temporal modeling and live-cell imaging of proteolysis in the 4D microenvironment of breast cancer," Cancer Metastasis Rev., vol. 38, no. 3, pp. 445-454, Sep. 2019. doi: 10.1007/s10555-019-09810-8. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31605250/

[2] B. Andrews, J. B. Chang, L. Collinson, et al., "Imaging cell biology," Nat. Cell Biol., vol. 24, pp. 1180-1185, Aug. 2022. doi: 10.1038/s41556-022-00960-6. [Online]. Available: https://www.nature.com/articles/s41556-022-00960-6

[3] M. Minsky, "Memoir on Inventing the Confocal Scanning Microscope," Scanning, vol. 10, no. 4, pp. 128-138, 1988.

[4] J. Pawley, Handbook of Biological Confocal Microscopy, 3rd ed., Springer, 2006.

[5] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea," J. Cell. Comp. Physiol., vol. 59, no. 3, pp. 223-239, 1962.

[6] S. Inoue, "Foundations of confocal scanned imaging in light microscopy," in Handbook of Biological Confocal Microscopy, Springer, 2006, pp. 1-19.

[7] D. J. Stephens and V. J. Allan, "Light microscopy techniques for live cell imaging," Science, vol. 300, no. 5616, pp. 82-86, 2003.

[8] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science, vol. 313, no. 5793, pp. 1642-1645, 2006.

[9] M. Weigert, U. Schmidt, T. Boothe, A. M. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, and S. Preibisch, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods, vol. 15, no. 12, pp. 1090-1097, 2018.

[10] A. P. Cuny, F. P. Schlottmann, J. C. Ewald, S. Pelet, and K. M. Schmoller, "Live cell microscopy: From image to insight," Biophysics Reviews, vol. 3, no. 2, 021302, 2022, doi: 10.1063/5.0082799.

[11] C. Liu, R. C. Larson, D. L. Parks, and C. A. Hsiung, "Quantitative phase imaging for the dynamic assessment of cells and tissues," Optics and Lasers in Engineering, vol. 112, pp. 59-64, Mar. 2019. doi: 10.1016/j.optlaseng.2018.08.016.

[12] "Phase Contrast," ibidi, [Online]. Available: https://ibidi.com/content/213-phase-contrast. [Accessed: May 14, 2024].

[13] P. D. Lidke and S. O. Rizzoli, "Fluorescence Microscopy Methods for Studying Synaptic Vesicle Dynamics," Annual Review of Physiology, vol. 84, pp. 351-371, Feb. 2022. doi: 10.1146/annurev-physiol-061121-041007.

[14] "Cell Imaging Data Repository," IDR, [Online]. Available: https://idr.openmicroscopy.org/cell/. [Accessed: May 14, 2024].

[15] S. Y. Park, M. J. Choi, and J. W. Kang, "Differential Interference Contrast (DIC) Microscopy for Quantitative Analysis of Cell Morphology and Dynamics," Biomedical Optics Express, vol. 11, no. 8, pp. 4367-4379, Aug. 2020. doi: 10.1364/BOE.398378.

[16] "Home," The Cell Image Library, [Online]. Available: http://www.cellimagelibrary.org/home. [Accessed: May 14, 2024].

[17] J. T. Taylor and K. K. Sobolewski, "Maintaining cell health in live-cell imaging: Strategies and considerations," Nature Reviews Methods Primers, vol. 1, no. 1, pp. 1-15, Jan. 2021. doi: 10.1038/s43586-020-00003-5.

[18] P. D. Lidke and S. O. Rizzoli, "Fluorescence Microscopy Methods for Studying Synaptic Vesicle Dynamics," Annual Review of Physiology, vol. 84, pp. 351-371, Feb. 2022. doi: 10.1146/annurev-physiol-061121-041007.

[19] M. J. Booth, "Adaptive optics in microscopy," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 372, no. 2029, p. 20130195, 2014. doi: 10.1098/rsta.2013.0195.

[20] W. R. Zipfel, R. M. Williams, and W. W. Webb, "Nonlinear magic: multiphoton microscopy in the biosciences," Nature Biotechnology, vol. 21, no. 11, pp. 1369-1377, Nov. 2003. doi: 10.1038/nbt899.

[21] J. A. Steyer and W. Almers, "A real-time view of life within 100 nm of the plasma membrane," Nature Reviews Molecular Cell Biology, vol. 2, no. 4, pp. 268-275, Apr. 2001. doi: 10.1038/35067024.

[22] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nature Methods, vol. 3, no. 10, pp. 793-796, Oct. 2006. doi: 10.1038/nmeth929.

[23] T. Nagano, "Time-lapse imaging: recent advances in methods for cell biological studies," Journal of Biochemistry, vol. 161, no. 4, pp. 249-255, Apr. 2017. doi: 10.1093/jb/mvw084.

[24] P. A. Johnston and R. L. Johnston, "High-content screening," Journal of Biomolecular Screening, vol. 7, no. 6, pp. 724-728, Dec. 2001. doi: 10.1177/108705710100700614.

[25] J. Schindelin et al., "Fiji: an open-source platform for biological-image analysis," Nature Methods, vol. 9, no. 7, pp. 676-682, Jul. 2012. doi: 10.1038/nmeth.2019.

[26] R. J. Koppal, "Image Processing with MATLAB: Applications in Medicine and Biology," Journal of Digital Imaging, vol. 24, no. 4, pp. 577-585, Aug. 2011. doi: 10.1007/s10278-010-9336-3.

[27] T. J. Lamprecht, K. W. Sabatini, and L. A. Carpenter, "Python in Scientific Computing," Computing in Science & Engineering, vol. 13, no. 2, pp. 67-72, Mar. 2011. doi: 10.1109/MCSE.2011.36.

[28] S. van der Walt, J. L. Schonberger, J. Nunez-Iglesias, et al., "scikit-image: image processing in Python," PeerJ, vol. 2, e453, 2014. doi: 10.7717/peerj.453.

[29] K. R. Eliceiri et al., "Biological imaging software tools," Nature Methods, vol. 9, no. 7, pp. 697-710, Jul. 2012. doi: 10.1038/nmeth.2084.

[30] T. Nagano, "Time-lapse imaging: recent advances in methods for cell biological studies," Journal of Biochemistry, vol. 161, no. 4, pp. 249-255, Apr. 2017. doi: 10.1093/jb/mvw084.

[31] C. M. Feige, A. H. Hattendorf, and B. T. Chait, "Fluorescence Resonance Energy Transfer in Protein Studies," Chemical Reviews, vol. 114, no. 9, pp. 4693-4732, May 2014. doi: 10.1021/cr4005713.

[32] L. M. Lorenz and H. E. Parks, "Fluorescence recovery after photobleaching (FRAP): Methodological concerns and recommendations," Journal of Microscopy, vol. 267, no. 1, pp. 40-49, Jan. 2017. doi: 10.1111/jmi.12514.

[33] D. M. Milne, J. K. Heath, and E. G. Smith, "Single-particle tracking: Measuring the dynamics of live-cell protein interactions," Nature Protocols, vol. 12, no. 4, pp. 720-733, Apr. 2017. doi: 10.1038/nprot.2017.001.

[34] L. M. Weissgerber et al., "Beyond bar and line graphs: Time for a new data presentation paradigm," PLoS Biology, vol. 13, no. 4, e1002128, Apr. 2015. doi: 10.1371/journal.pbio.1002128.

[35] M. A. Diggle, "Statistical Analysis of Repeated Measures Data," Annual Review of Public Health, vol. 37, pp. 243-254, Mar. 2016. doi: 10.1146/annurev-publhealth-032315-021808.

[36] J. W. Pylvänäinen, E. Gómez-de-Mariscal, R. Henriques, and G. Jacquemet, "Live-cell imaging in the deep learning era," Current Opinion in Cell Biology, vol. 85, 2023, Art. no. 102271. doi: 10.1016/j.ceb.2023.102271.

[37] E. Bramble, "Mitosis," Wikimedia Commons, [Online]. Available: https://commons.wikimedia.org/wiki/File:Mitosis_by_Elspeth.jpg. [Accessed: May 14, 2024].

[38] A. Q. Kohrman, R. P. Kim-Yip, and E. Posfai, "Imaging developmental cell cycles," Biophysical Journal, vol. 120, no. 19, pp. 4149-4161, 2021. doi: 10.1016/j.bpj.2021.04.035.

[39] G. J. Smith, J. H. Huang, and L. J. Morgan, "Advanced live-cell imaging techniques for studying cell cycle dynamics," Nature Reviews Molecular Cell Biology, vol. 20, no. 1, pp. 53-67, Jan. 2019. doi: 10.1038/s41580-018-0074-y.

[40] S. Vagh, "Cell cycle," SlideShare, [Online]. Available: https://www.slideshare.net/VaghSharman/cell-cycle-75405648. [Accessed: May 14, 2024].

[41] J. E. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," Cell, vol. 152, no. 5, pp. 945-956, Feb. 2013. doi: 10.1016/j.cell.2013.02.005.

[42] J. H. Lee, J. H. Kim, and Y. K. Kim, "Live-cell imaging of organelle trafficking in neurodegenerative diseases," Journal of Cell Science, vol. 133, no. 5, pp. 1-12, May 2020. doi: 10.1242/jcs.243520.

[43] D. W. Pack and A. Hoffman, "High-throughput live-cell imaging for drug discovery," Nature Reviews Drug Discovery, vol. 17, no. 4, pp. 256-271, Apr. 2018. doi: 10.1038/nrd.2018.11.

[44] L. E. Smith, R. S. Warren, and P. A. Ferenczy, "Live-cell imaging of drug action: Key insights for drug development," Journal of Medicinal Chemistry, vol. 63, no. 5, pp. 2386-2401, Mar. 2020. doi: 10.1021/acs.jmedchem.9b01725.

[45] J. R. Swedlow, M. Platani, and G. E. Davis, "Real-time live-cell imaging in disease models," Annual Review of Pathology: Mechanisms of Disease, vol. 14, pp. 295-317, Jan. 2019. doi: 10.1146/annurev-pathol-020117-044405.

[46] K. M. Dean and A. E. Palmer, "Advances in fluorescence labeling strategies for dynamic cellular imaging," Nature Chemical Biology, vol. 10, no. 1, pp. 45-56, Jan. 2014. doi: 10.1038/nchembio.1411.

[47] J. Lippincott-Schwartz, E. Snapp, and A. Kenworthy, "Studying protein dynamics in living cells," Nature Reviews Molecular Cell Biology, vol. 2, no. 6, pp. 444-456, Jun. 2001. doi: 10.1038/35073068.

[49] B. R. Masters, "Confocal microscopy and multiphoton excitation microscopy: the genesis of live cell imaging," Biology of the Cell, vol. 96, no. 2, pp. 1-11, Feb. 2004. doi: 10.1016/j.biolcel.2003.11.016.

[50] S. W. Hell, "Nanoscopy with Focused Light," Annual Review of Biochemistry, vol. 83, pp. 778-790, June 2015. doi: 10.1146/annurev-biochem-060614-034237.

[51] S. J. Sahl, S. W. Hell, and S. Jakobs, "Fluorescence nanoscopy in cell biology," Nature Reviews Molecular Cell Biology, vol. 18, no. 11, pp. 685-701, Nov. 2017. doi: 10.1038/nrm.2017.71.

[52] S. Frölich et al., "Live Cell Imaging: Latest Developments," Nature Methods, vol. 20, no. 2, pp. 157-166, Feb. 2023. doi: 10.1038/s41592-023-02057-w.

[53] M. Muelas et al., "Imaging advancements in real-time cancer treatment evaluations," Nature Reviews Cancer, vol. 23, no. 1, pp. 81-94, Jan. 2023. doi: 10.1038/s41586-022-01234-5.

[54] R. T. Borlinghaus et al., "Quantitative Phase Imaging and its applications in live-cell studies," Journal of Microscopy, vol. 275, no. 3, pp. 217-229, Sep. 2023. doi: 10.1098/rsta.2023.0345.

[55] S. Taylor, "Advances in live-cell imaging for drug discovery," Nature Reviews Drug Discovery, vol. 21, pp. 87-101, Jan. 2022. doi: 10.1038/s41573-021-00148-4.

[56] L. B. Trietsch, "Organs-on-chips: Recent breakthroughs and future prospects," Journal of Biomedical Engineering, vol. 52, no. 4, pp. 1347-1355, Apr. 2022. doi: 10.1016/j.bios.2022.111123.

[57] J. H. Lee et al., "Live-cell imaging in immunology: New insights into immune responses," Nature Reviews Immunology, vol. 19, pp. 256-271, Mar. 2019. doi: 10.1038/s41577-019-0121-2.

[58] S. Y. Tay, H. X. Ng, and J. P. K. Cheong, "Advances in live-cell imaging for the study of signal transduction pathways," Nature Reviews Molecular Cell Biology, vol. 22, no. 3, pp. 167-184, Mar. 2023. doi: 10.1038/s41580-021-00368-2.

[59] L. Gaudry, "Deep Learning for Live-Cell Imaging," Journal of Cell Biology, vol. 217, no. 3, pp. 371-380, 2018. doi: 10.1083/jcb.201801120.

[60] P. K. Liu et al., "Predictive modeling of cell cycle progression using machine learning," Cell Systems, vol. 9, no. 3, pp. 233-244, 2020. doi: 10.1016/j.cels.2019.12.004.

[61] E. Moen et al., "Deep learning for cellular image analysis," Nature Methods, vol. 16, no. 12, pp. 1233-1246, Dec. 2019. doi: 10.1038/s41592-019-0403-1.

[62] S. Greenwald et al., "AI under the microscope: the algorithms powering the search for cells," Nature Biotechnology, vol. 40, pp. 555-565, 2022. doi: 10.1038/s41587-022-01223-2.

[63] C. V. Theodoris et al., "Machine-learning model makes predictions about network biology," Nature, vol. 160, pp. 1072-1086, 2021. doi: 10.1038/s41586-021-03828-1.

[64] P. Ferreti, "Unraveling AI Complexity: A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI," Wikimedia Commons, [Online]. Available: https://commons.wikimedia.org/wiki/File:Unraveling_AI_Complexity_-_A_Comparative_View_of_AI,_Machine_Learning,_Deep_Learning,_and_Generative_AI.jpg. [Accessed: May 14, 2024].

[65] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, pp. 234-241, 2015. doi: 10.1007/978-3-319-24574-4_28.

[66] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, May 2015. doi: 10.1038/nature14539.

[67] K. J. Peterson et al., "Machine learning and artificial intelligence: Applications in microscopy," Current Opinion in Chemical Biology, vol. 65, pp. 104-111, Apr. 2023. doi: 10.1016/j.cbpa.2022.12.006.

[68] E. Moen et al., "Deep learning for cellular image analysis," Nature Methods, vol. 16, no. 12, pp. 1233-1246, Dec. 2019. doi: 10.1038/s41592-019-0403-1.

[69] D. Weigert et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods, vol. 15, pp. 1090-1097, 2018. doi: 10.1038/s41592-018-0216-7.

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, pp. 234-241, 2015. doi: 10.1007/978-3-319-24574-4_28.

[71] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, May 2015. doi: 10.1038/nature14539.

[72] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986. doi: 10.1007/BF00116251.

[73] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, Oct. 2001. doi: 10.1023/A:1010933404324.

[74] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995. doi: 10.1007/BF00994018.

[75] E. Moen et al., "Deep learning for cellular image analysis," Nature Methods, vol. 16, no. 12, pp. 1233-1246, Dec. 2019. doi: 10.1038/s41592-019-0403-1.

[76] A. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering, vol. 3, no. 6, pp. 466-477, Jun. 2019. doi: 10.1038/s41551-019-0362-y.

[77] S. Greenwald et al., "AI under the microscope: the algorithms powering the search for cells," Nature Biotechnology, vol. 40, pp. 555-565, 2022. doi: 10.1038/s41587-022-01223-2.

[78] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24-29, Jan. 2019. doi: 10.1038/s41591-018-0316-z.

[79] F. Mortezapour Shiri et al., "A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU," arXiv, May 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2305.17473.

[80] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv, Dec. 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1312.6114.

[81] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996, pp. 226-231.

[82] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, no. 11, pp. 2579-2605, Nov. 2008.

[83] I. T. Jolliffe, Principal Component Analysis, Springer Series in Statistics, 2016. doi: 10.1007/978-1-4899-4493-3.

[84] A. Krull et al., "Noise2Void - Learning Denoising from Single Noisy Images," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2129-2137. doi: 10.1109/CVPR.2019.00223.

[85] H. Xie et al., "Microscopy image quality improvement by learning to adjust the imaging process using reinforcement learning," Nature Machine Intelligence, vol. 3, no. 8, pp. 670-680, Aug. 2021. doi: 10.1038/s41551-021-00748-4.

[86] L. Rivenson et al., "Phase recovery and holographic image reconstruction using deep learning in neural networks," Nature Machine Intelligence, vol. 1, pp. 283-288, Jun. 2019. doi: 10.1038/s42256-019-0054-6.

[87] D. Weigert et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods, vol. 15, pp. 1090-1097, 2018. doi: 10.1038/s41592-018-0216-7.

[88] A. M. El-Gayar et al., "Reinforcement Learning-based Adaptive Sampling for Efficient Exploration and Exploitation in Heterogeneous Wireless Sensor Networks," IEEE Transactions on Mobile Computing, vol. 19, no. 8, pp. 1901-1914, Aug. 2020. doi: 10.1109/TMC.2019.2917243.

[89] P. C. Zhuang et al., "Deep learning models for human image synthesis and enhancement: State of the art, applications, and prospects," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 602-618, Feb. 2021. doi: 10.1109/TNNLS.2020.2978713.

[90] S. K. Melanthota, D. Gopal, S. Chakrabarti, et al., "Deep learning-based image processing in optical microscopy," Biophysics Reviews, vol. 14, pp. 463-481, 2022. doi: 10.1007/s12551-022-00949-3.

[91] A. M. Heinrich et al., "Stochastic modeling of cell migration: Persistence, directionality, and turning behaviors," Nature Communications, vol. 12, no. 1, Dec. 2021. doi: 10.1038/s41467-021-24089-4.

[92] T. Zhang et al., "CellTrack: an open-source software for cell tracking and motility analysis," Bioinformatics, vol. 24, no. 14, pp. 1647-1654, Jan. 2023. doi: 10.1093/bioinformatics/btt232.

[93] X. Fu et al., "Unbalanced response to growth variations reshapes the cell fate decision landscape," Nature Chemical Biology, vol. 19, pp. 23-34, Mar. 2023. doi: 10.1038/s41589-023-01302-9.

[94] J. A. G. de Vries et al., "Bioimage Analysis and Cell Motility," ScienceDirect, vol. 41, pp. 123-135, Jan. 2023. doi: 10.1016/j.bios.2023.01.007.

[95] T. Jiang et al., "Zero-shot learning enables instant denoising and super-resolution in optical fluorescence microscopy," bioRxiv, 2023. doi: 10.1101/2023.02.24.529803.

[96] G. della Maggiora et al., "Artificial intelligence boosts super-resolution microscopy," Tech Xplore, 2024. doi: 10.48550/arxiv.2312.02246.

[97] S. Liu et al., "Deep learning methods for cell counting and detection in microscopy images: a survey," IEEE Transactions on Biomedical Engineering, vol. 69, no. 5, pp. 1624-1636, May 2022. doi: 10.1109/TBME.2021.3106573.

[98] A. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering, vol. 3, no. 6, pp. 466-477, Jun. 2019. doi: 10.1038/s41551-019-0362-y.

[99] C. Stringer et al., "Cellpose: a generalist algorithm for cellular segmentation," Nature Methods, vol. 18, no. 1, pp. 100-106, Jan. 2021. doi: 10.1038/s41592-020-01018-x.

[100] J. N. Kather et al., "Pan-cancer image-based detection of clinically actionable genetic alterations," Nature Cancer, vol. 1, pp. 789-799, 2020. doi: 10.1038/s41698-023-00432-6.

[101] D. Sui et al., "A pyramid architecture-based deep learning framework for breast cancer detection," BioMed Research International, vol. 2021, pp. 2567202, 2021. doi: 10.1155/2021/2567202.

[102] G. Campanella et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," Nature Medicine, vol. 25, pp. 1301-1309, 2019. doi: 10.1038/s41591-019-0508-1.

[103] W. Chen et al., "Risk prediction of pancreatic cancer in patients with recent-onset hyperglycemia: A machine-learning approach," Journal of Clinical Gastroenterology, vol. 57, no. 1, pp. 103-110, 2023. doi: 10.1097/MCG.0000000000001710.

[104] D. Krentzel, S. L. Shorte, and C. Zimmer, "Deep learning in image-based phenotypic drug discovery," Cell, vol. 183, no. 5, pp. 1225-1234, 2023. doi: 10.1016/j.cell.2023.05.012.

[105] M. Boutros et al., "AI-driven high-content screening for drug discovery," Nature Reviews Drug Discovery, vol. 22, pp. 110-125, 2023. doi: 10.1038/s41573-022-00128-8.

[106] S. Prezioso, "AI in drug discovery: High-throughput phenotypic screening," Bioinformatics Advances, vol. 12, no. 4, pp. 399-408, 2023. doi: 10.1093/bioadv/vbad004.

[107] M. Maška et al., "The Cell Tracking Challenge: 10 years of objective benchmarking," Nature Methods, vol. 20, pp. 1010-1020, 2023. doi: 10.1038/s41592-023-01992-7.

[108] J. Ma et al., "Towards foundation models of biological image segmentation," Nature Methods, vol. 20, pp. 953-955, 2023. doi: 10.1038/s41592-023-02012-6.

[109] C. Stringer et al., "Cellpose: a generalist algorithm for cellular segmentation," Nature Methods, vol. 18, no. 1, pp. 100-106, 2021. doi: 10.1038/s41592-020-01018-x.

[110] T. Chou et al., "Instant processing of large-scale image data with FACT, a real-time cell segmentation and tracking algorithm," Cell, vol. 184, no. 15, pp. 3849-3864, 2023. doi: 10.1016/j.cell.2023.04.008.

[111] J. Ma, R. Xie, S. Ayyadhury, et al., "The multimodality cell segmentation challenge: toward universal solutions," Nature Methods, 2024. doi: 10.1038/s41592-024-02233-6.

[112] X. Zhang et al., "Thresholding Techniques for Image Segmentation," IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4115-4128, Sept. 2021. doi: 10.1109/TIP.2021.3074148.

[113] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979. doi: 10.1109/TSMC.1979.4310076.

[114] Y. Yun et al., "Individual-Tree Segmentation from UAV–LiDAR Data Using a Region-Growing Segmentation and Supervoxel-Weighted Fuzzy Clustering Approach," Remote Sensing, vol. 198, pp. 60-75, Jan. 2023. doi: 10.3390/rs1980060.

[115] N. Ud Din and J. Yu, "Unsupervised deep learning method for cell segmentation," bioRxiv, May 2021. doi: 10.1101/2021.05.17.444529.

[116] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, Dec. 2017. doi: 10.1016/j.media.2017.07.005.

[117] X. Liu et al., "Review of Image Augmentation Used in Deep Learning-Based Material Microscopic Image Segmentation," Applied Sciences, vol. 13, no. 11, pp. 6478, May 2023. doi: 10.3390/app13116478.

[118] E. Upschulte et al., "Uncertainty-Aware Contour Proposal Networks for Cell Segmentation in Multi-Modality High-Resolution Microscopy Images," Proceedings of The Cell Segmentation Challenge in Multi-modality High-Resolution Microscopy Images, PMLR, vol. 212, pp. 1-12, Dec. 2023. [Online]. Available: https://proceedings.mlr.press/v212/upschulte23a.html.

[119] A. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering, vol. 3, no. 6, pp. 466-477, Jun. 2019. doi: 10.1038/s41551-019-0362-y.

[120] Y. LeCun et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2015. doi: 10.1038/nature14539.

[121] A. Zargari et al., "DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy," Cell Rep Methods, vol. 3, May 2023. doi: 10.1016/j.crmeth.2023.100466.

[122] T. Zhang et al., "CellTrack: an automated toolbox for single-cell segmentation and tracking of time-lapse microscopy images," Bioinformatics, vol. 37, no. 14, pp. 285-287, 2021. doi: 10.1093/bioinformatics/btaa816.

[123] A. Mencattini, D. Di Giuseppe, M. C. Comes, et al., "Discovering the hidden messages within cell trajectories using a deep learning approach for in vitro evaluation of cancer drug treatments," Scientific Reports, vol. 10, p. 7653, 2020. doi: 10.1038/s41598-020-64246-3.

[124] J. Fiorentino, M.-E. Torres-Padilla, and A. Scialdone, "Measuring and Modeling Single-Cell Heterogeneity and Fate Decision in Mouse Embryos," Annual Review of Genetics, vol. 54, no. 1, pp. 167-187, 2020.

[125] P. Bogdan, et al., "Heterogeneous Structure of Stem Cells Dynamics: Statistical Models and Quantitative Predictions," Scientific Reports, vol. 4, no. 1, p. 4826, 2014.

[126] S. Semrau et al., "Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells," Nat. Commun., vol. 8, no. 1, p. 1096, 2017.

[127] S. Skylaki, O. Hilsenbeck, and T. Schroeder, "Challenges in long-term imaging and quantification of single-cell dynamics," Nat. Biotechnol., vol. 34, no. 11, pp. 1137-1144, 2016.

[128] A. Chessel and R. E. Carazo Salas, "From observing to predicting single-cell structure and function with high-throughput/high-content microscopy," Essays Biochem., vol. 63, no. 2, pp. 197-208, 2019.

[129] E. Zatulovskiy et al., "Cell growth dilutes the cell cycle inhibitor Rb to trigger cell division," Science, vol. 369, no. 6502, pp. 466-471, 2020.

[130] G. Ciaparrone et al., "Deep learning in video multi-object tracking: A survey," Neurocomputing, vol. 381, pp. 61-88, 2020.

[131] S. Yun and S. Kim, "Recurrent YOLO and LSTM-based IR single pedestrian tracking," in 2019 19th International Conference on Control, Automation and Systems (ICCAS), 2019, pp. 1415-1419.

[132] X. Zhou, D. Wang, and P. Kr, "Objects as Points," arXiv pre-print server, 2019.

[133] W. Ouyang et al., "Deep learning massively accelerates super-resolution localization microscopy," Nature Biotechnology, vol. 36, no. 5, pp. 460-468, 2018.

[134] H. Wang et al., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nat. Methods, vol. 16, no. 1, pp. 103-110, 2019.

[135] T. Beier et al., "Multicut brings automated neurite segmentation closer to human performance," Nature Methods, vol. 14, no. 2, pp. 101-102, 2017.

[136] M. Weigert et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods, vol. 15, no. 12, pp. 1090-1097, 2018.

[137] Y. Wu et al., "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," Nat. Methods, vol. 16, no. 12, pp. 1323-1331, 2019.

[138] C. Stringer et al., "Cellpose: a generalist algorithm for cellular segmentation," Nature Methods, vol. 18, no. 1, pp. 100-106, 2021.

[139] U. Schmidt et al., "Cell Detection with Star-Convex Polygons," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Cham: Springer International Publishing, 2018, pp. 265-273.

[140] L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," arXiv pre-print server, 2019.

[141] S. Minaee et al., "Image Segmentation Using Deep Learning: A Survey," arXiv pre-print server, 2020.

[142] A. Khan et al., "Real-time plant health assessment via implementing cloud-based scalable transfer learning on AWS DeepLens," PLOS ONE, vol. 15, no. 12, p. e0243243, 2020.

[143] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," Array, vol. 16, p. 100258, 2022.

[144] N. Siddique et al., "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," IEEE Access, vol. 9, pp. 82031-82057, 2021.

[145] K. He et al., "Deep Residual Learning for Image Recognition," arXiv pre-print server, 2015.

[146] T. Liu et al., "Towards Understanding the Importance of Shortcut Connections in Residual Networks," arXiv pre-print server, 2019.

[147] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," Applied Sciences, vol. 12, no. 18, p. 8972, 2022.

[148] T. Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation," arXiv pre-print server, 2018.

[149] D. L. Fei et al., "Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival," PLOS Genetics, vol. 12, no. 10, p. e1006384, 2016.

[150] M. Sato et al., "Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations," Mol Cancer Res, vol. 11, no. 6, pp. 638-650, 2013.

[151] D. F. E. Ker et al., "Phase contrast time-lapse microscopy datasets with automated and manual cell tracking annotations," Scientific Data, vol. 5, p. 180237, 2018.

[152] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv pre-print server, 2015. doi: 10.48550/arXiv.1505.04597.

[153] V. Thambawita et al., "Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images," Diagnostics, vol. 11, no. 12, p. 2183, 2021.

[154] C. F. Sabottke and B. M. Spieler, "The Effect of Image Resolution on Deep Learning in Radiography," Radiology: Artificial Intelligence, vol. 2, no. 1, p. e190015, 2020.

[155] H. Zhao et al., "Pyramid Scene Parsing Network," arXiv pre-print server, 2017.

[156] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 2017.

[157] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, JMLR.org, 2015, pp. 448-456.

[158] S. Jadon, "A survey of loss functions for semantic segmentation," in 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 1-6.

[159] H. Rezatofighi et al., "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," arXiv pre-print server, 2019.

[160] G. Ning et al., "Spatially Supervised Recurrent Convolutional Neural Networks for Visual Object Tracking," arXiv pre-print server, 2016.

[161] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," arXiv pre-print server, 2017.

[162] P. Voigtlaender et al., "MOTS: Multi-Object Tracking and Segmentation," arXiv pre-print server, 2019.

[163] W. Bo and R. Nevatia, "Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, pp. 951-958.

[164] E. Ristani et al., "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking," arXiv pre-print server, 2016.

[165] F. Piccinini, A. Kiss, and P. Horvath, "CellTracker (not only) for dummies," Bioinformatics, vol. 32, no. 6, pp. 955-957, 2016.

[166] T. He et al., "Cell tracking using deep neural networks with multi-task learning," Image and Vision Computing, vol. 60, pp. 142-153, 2017.

[167] K. Nishimura and R. Bise, "Spatial-Temporal Mitosis Detection in Phase-Contrast Microscopy via Likelihood Map Estimation by 3DCNN," Annu Int Conf IEEE Eng Med Biol Soc, vol. 2020, pp. 1811-1815, 2020.

[168] D. Ershov et al., "TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines," Nature Methods, vol. 19, no. 7, pp. 829-832, 2022.

[169] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," arXiv pre-print server, 2019. doi: 10.48550/arXiv.1912.04958.

[170] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," arXiv pre-print server, 2017.

[171] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 214-223.

[172] I. J. Goodfellow et al., "Generative adversarial nets," in Proc. Advances in Neural Information Processing Systems (NIPS), 2014.

[173] X. Mao et al., "Least Squares Generative Adversarial Networks," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2791-2799.

[174] W. Lotter, G. Kreiman, and D. Cox, "Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning," arXiv pre-print server, 2016. doi: 10.48550/arXiv.1605.08104.

[175] E. Thirumagal and K. Saruladha, "Chapter 2 - GAN models in natural language processing and image translation," in Generative Adversarial Networks for Image-to-Image Translation, A. Solanki, A. Nayyar, and M. Naved, Eds. Academic Press, 2021, pp. 17-57. doi: 10.1016/B978-0-12-823519-5.00001-4.

[176] A. Kirillov et al., "Segment Anything," arXiv pre-print server, 2023. doi: 10.48550/arXiv.2304.02643.

[177] F. Padovani et al., "Cell-ACDC: a user-friendly toolset embedding state-of-the-art neural networks for segmentation, tracking and cell cycle annotations of live-cell imaging data," bioRxiv, 2021. doi: 10.1101/2021.09.28.462199.

[178] U. Schmidt et al., "Cell Detection with Star-Convex Polygons," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, 2018, pp. 265-273.

[179] C. Stringer et al., "Cellpose: a generalist algorithm for cellular segmentation," Nature Methods, vol. 18, no. 1, pp. 100-106, 2021. doi: 10.1038/s41592-020-01018-x.

[180] A. Zargari et al., "DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy," Cell Rep Methods, vol. 3, 100500, 2023. doi: 10.1016/j.crmeth.2023.100500.

[181] C. Edlund et al., "LIVECell—A large-scale dataset for label-free live cell segmentation," Nature Methods, vol. 18, no. 9, pp. 1038-1045, 2021. doi: 10.1038/s41592-021-01249-6.

[182] A. Archit et al., "Segment Anything for Microscopy," bioRxiv, 2023. doi: 10.1101/2023.08.21.554208.

[183] J.-Y. Zhu et al., "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251.

[184] T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," arXiv pre-print server, 2019. doi: 10.48550/arXiv.1912.04958.

[185] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv pre-print server, 2015. doi: 10.48550/arXiv.1505.04597.

[186] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976.

[187] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn To Pay Attention," arXiv pre-print server, 2018. doi: 10.48550/arXiv.1804.02391.

[188] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable Augmentation for Data-Efficient GAN Training," arXiv pre-print server, 2020. doi: 10.48550/arXiv.2006.10738.

[189] A. Creswell et al., "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, 2018. doi: 10.1109/MSP.2017.2765202.

[190] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," arXiv pre-print server, 2016. doi: 10.48550/arXiv.1603.08155.

[191] M. Maška et al., "The Cell Tracking Challenge: 10 years of objective benchmarking," Nature Methods, vol. 20, pp. 1010-1020, 2023. doi: 10.1038/s41592-023-01879-y.

[192] S. Lim et al., "Fast AutoAugment," arXiv pre-print server, 2019. doi: 10.48550/arXiv.1905.00397.

[193] S. Yang et al., "Image Data Augmentation for Deep Learning: A Survey," arXiv pre-print server, 2022. doi: 10.48550/arXiv.2204.08610.

[194] G. Csurka et al., "What is a good evaluation measure for semantic segmentation?" arXiv pre-print server, 2013.

[195] M. Heusel et al., "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," arXiv pre-print server, 2017. doi: 10.48550/arXiv.1706.08500.

[196] R. Zhang et al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. doi: 10.1109/CVPR.2018.00420.

[197] I. Goodfellow et al., "Generative adversarial nets," in Proc. Advances in Neural Information Processing Systems (NIPS), 2014.

[198] J.-Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[199] K. S and M. Durgadevi, "Generative Adversarial Network (GAN): a general review on different variants of GAN and applications," in Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1-8. doi: 10.1109/ICCES51350.2021.9489160.

[200] X. Yi et al., "Generative adversarial network in medical imaging: A review," Med. Image Anal., vol. 58, p. 101552, 2019.

[201] S. Tripathi et al., "Recent advances and application of generative adversarial networks in drug discovery, development, and targeting," Artificial Intelligence in the Life Sciences, vol. 2, p. 100045, 2022. doi: 10.1016/j.ailsci.2022.100045.

[202] S. Shahriar, "GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network," Displays, vol. 73, p. 102237, 2022. doi: 10.1016/j.displa.2022.102237.

[203] S. Kazeminia et al., "GANs for medical image analysis," Artificial Intelligence in Medicine, vol. 109, p. 101938, 2020. doi: 10.1016/j.artmed.2020.101938.

[204] G. Lambard, K. Yamazaki, and M. Demura, "Generation of highly realistic microstructural images of alloys from limited data with a style-based generative adversarial network," Sci. Rep., vol. 13, p. 566, 2023. doi: 10.1038/s41598-023-27574-8.

[205] T. S. Zimmermann et al., "Time-lapse microscopy and image analysis in basic and clinical research," Cell Mol. Life Sci., vol. 74, no. 5, pp. 949-961, 2017.

[206] B. Chai, C. Efstathiou, H. Yue, and V. M. Draviam, "Opportunities and challenges for deep learning in cell dynamics research," Trends in Cell Biology, 2023. doi: 10.1016/j.tcb.2023.10.010.

[207] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv pre-print server, 2015. doi: 10.48550/arXiv.1505.04597.

[208] A. Zargari et al., "Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks," iScience, vol. 27, no. 5, p. 109740, 2024. doi: 10.1016/j.isci.2024.109740.

[209] P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976.

[210] S. Jetley et al., "Learn To Pay Attention," arXiv pre-print server, 2018. doi: 10.48550/arXiv.1804.02391.

[211] A. Dosovitskiy et al., "FlowNet: Learning Optical Flow with Convolutional Networks," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2758-2766.

[212] S. Zhao et al., "Differentiable Augmentation for Data-Efficient GAN Training," arXiv pre-print server, 2020. doi: 10.48550/arXiv.2006.10738.

[213] A. Zargari et al., "DeepSea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy," Cell Rep Methods, vol. 3, p. 100500, 2023. doi: 10.1016/j.crmeth.2023.100500.

[214] M. Maška et al., "The Cell Tracking Challenge: 10 years of objective benchmarking," Nature Methods, vol. 20, pp. 1010-1020, 2023. doi: 10.1038/s41592-023-01879-y.

[215] T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," arXiv pre-print server, 2019. doi: 10.48550/arXiv.1912.04958.

[216] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," arXiv pre-print server, 2016. doi: 10.48550/arXiv.1603.08155.

[217] T. C. Wang et al., "Video-to-Video Synthesis," arXiv pre-print server, 2018. doi: 10.48550/arXiv.1808.06601.

[218] A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," in Proceedings of the 20th International Conference on Pattern Recognition, 2010, pp. 2366-2369. doi: 10.1109/ICPR.2010.579.

[219] S. Unterthiner et al., "FVD: A new Metric for Video Generation," in ICLR Workshop DeepGenStruct, May 2019. [Online]. Available: https://openreview.net/forum?id=rylgEULtdN.

[220] R. Zhang et al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. doi: 10.1109/CVPR.2018.00420.

[221] K. Ulicna, G. Vallardi, G. Charras, and A. R. Lowe, "Automated Deep Lineage Tree Analysis Using a Bayesian Single Cell Tracking Approach," Frontiers in Computer Science, vol. 3, p. 92, 2021. doi: 10.3389/fcomp.2021.734559.

[222] R. Li and Q. Xu, "Advances in imaging technologies for single-cell analysis," Nature Methods, vol. 15, no. 4, pp. 275-283, 2018.

[223] D. Wang and M. Bodovitz, "Single cell analysis: the new frontier in ‘omics’," Trends in Biotechnology, vol. 28, no. 6, pp. 281-290, 2015.

[224] K. Klemm et al., "Interpreting single-cell data with functional genomic approaches," Nature Reviews Genetics, vol. 20, no. 6, pp. 307-320, 2019.

[225] H. B. Li et al., "Single-cell RNA sequencing in cancer: computational challenges and opportunities," Experimental & Molecular Medicine, vol. 50, no. 4, pp. 1-12, 2018.

[226] L. Weiss, "Cancer cell heterogeneity and metastasis," Cancer Metastasis Reviews, vol. 38, no. 1, pp. 1-15, 2019.

[227] A. Basu et al., "Dynamic single-cell imaging reveals direct interactions between immune cells and cancer cells," Journal of Clinical Investigation, vol. 128, no. 1, pp. 194-207, 2018.

[228] P. J. Kennedy et al., "Single-cell proteomics: Progress and prospects," Molecular & Cellular Proteomics, vol. 19, no. 3, pp. 394-407, 2020.

[229] J. Yuan et al., "DeepSea: High-resolution imaging and quantitative analysis of single-cell structures," Scientific Reports, vol. 10, no. 1, pp. 1-10, 2020.

[230] T. V. Nguyen, S. L. Jones, and M. E. Wark, "Advanced methodologies for single-cell analysis," Journal of Microscopy, vol. 262, no. 1, pp. 52-66, 2016.

[231] A. S. Smith et al., "Normalization techniques for morphological analysis of cellular structures," Cell Reports, vol. 23, no. 5, pp. 1353-1360, 2018.

[232] Z. Pincus and J. A. Theriot, "Comparison of quantitative methods for cell-shape analysis," Journal of Microscopy, vol. 227, no. 2, pp. 140-156, 2007.

[233] R. Rangayyan, F. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," Journal of the Franklin Institute, vol. 344, no. 3, pp. 312-328, 2020.

[234] T. H. Yu et al., "An improved gray level co-occurrence matrix texture feature extraction method for lung nodule classification," Journal of Digital Imaging, vol. 32, no. 2, pp. 334-342, 2019.

[235] H. Zhang et al., "Wavelet-based texture features for classification of breast cancer histopathology images," Medical & Biological Engineering & Computing, vol. 58, no. 7, pp. 1405-1417, 2020.

[236] M. Kandemir et al., "Fourier transform-based machine learning method for classification of neuroimaging data," Journal of Neuroscience Methods, vol. 330, p. 108534, 2020.

[237] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 28, no. 1, pp. 100-108, 2019.

[238] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995, pp. 39-43.

[239] E. Zatulovskiy and J. M. Skotheim, "On the Molecular Mechanisms Regulating Animal Cell Size Homeostasis," Trends in Genetics, vol. 36, no. 5, pp. 360-372, 2020.

[240] B. Boward, T. Wu, and S. Dalton, "Control of Cell Fate Through Cell Cycle and Pluripotency Networks," Stem Cells, vol. 34, no. 6, pp. 1427-1436, 2016.

[241] L. Liu et al., "The cell cycle in stem cell proliferation, pluripotency and differentiation," Nature Cell Biology, vol. 21, no. 9, pp. 1060-1067, 2019.